(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number
WO 01/27632 A2

(51) International Patent Classification[7]: G01N 33/68

(21) International Application Number: PCT/IB00/01407

(22) International Filing Date: 2 October 2000 (02.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **CAMBRIDGE DRUG DISCOVERY, LTD.** [GB/GB]; Cambridge Science Park, Milton Road, Cambridge CB4 0FG (GB).

(71) Applicant *(for BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG only)*: **WILLIAMS, Kathleen, M.** [US/US]; 10 Nathaniel Guild Road, Sharon, MA 02067 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WENHAM, Dean** [GB/GB]; Flt 5 Riverside, London Road, St. Ives. PE27 4GU (GB). **PACKER, Jeremy, Charles** [GB/GB]; 34 Cloverfield Drive, Soham ELY CB7 (GB).

(74) Common Representative: **WILLIAMS, Kathleen, M.**; Palmer & Dodge LLP, One Beacon Street, Boston, MA 02108 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a).*

— *Without international search report and to be republished upon receipt of that report.*

— *Without classification; title and abstract not checked by the International Searching Authority.*

— *With an indication in relation to a priority claim considered not to have been made.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/27632 A2

(54) Title: METHOD OF PREDICTING MUTATIONS

(57) Abstract: The invention relates to methods of predicting mutations that alter the activity of a receptor in a desired manner. The methods utilize multiple sequence alignment and phylogenetic profiling to identify the relatives of a given receptor that are most likely to provide useful data allowing prediction of sites to mutate in the given receptor. The methods provided are applicable to any type of receptor, and are particularly well suited for predicting sites to mutate in order to alter the activities of the so-called orphan receptors, for which no agonists are known.

METHOD OF PREDICTING MUTATIONS

This application claims the priority of U.K. Application No. 9922986.6, filed September 29, 1999, which is incorporated herein in its entirety.

TECHNICAL AREA OF THE INVENTION

The general technical area of the invention is receptor mutagenesis. In particular, the invention relates to methods of predicting specific sites to mutate in a cellular receptor in order to achieve a desired functional alteration. More particularly, the invention relates to the prediction of cellular receptor mutations that induce the receptor to constitutively activate its downstream signaling activities.

BACKGROUND OF THE INVENTION

The control of intracellular events is affected by extracellular signals through the activation of cellular receptors. Generally, the interaction of a cellular receptor with its cognate ligand results in the generation of an intracellular signal that affects the activity of one or more downstream signal transduction pathways, culminating in an effect on cellular metabolism or gene expression.

Receptors may be transmembrane proteins anchored in the plasma membrane. Transmembrane receptors interact with extracellular ligands to transduce a signal. Alternatively, receptors may be intracellular proteins that interact with ligands that pass through the plasma membrane. Intracellular receptor proteins tend to have the ability, upon activation, to translocate from the cytoplasm to the nucleus, where they often have a direct effect on gene expression.

The signaling molecules that interact with cellular receptors include hormones and growth factors, ions, neurotransmitters, cytokines and other molecules involved in immune regulation.

The specific mechanisms of signal transduction are not fully understood, but it is clear that a number of disease states involve the inappropriate activation or inactivation of receptor signaling pathways. This inappropriate regulation occurs through various means, including overproduction of growth factors or hormones, mutation of receptors themselves, and disturbances affecting other proteins, such as kinases and phosphatases, that participate in the downstream signal transduction pathways. There is a widely held desire in the scientific and

medical communities to identify agents that inhibit or activate cellular receptors, in order to control diseases involving receptor or receptor signal transduction pathway activation.

Traditionally, receptors for which a ligand is known have been studied using that natural ligand. Agents may be identified for example, that inhibit the recognition of the ligand by the receptor or that render the ligand unable to bind the receptor. One major drawback of these approaches is that they tend to only be useful for the identification of agents that work upstream of receptor activation. Such agents are not likely to be useful for treating conditions involving disturbances in the signal transduction pathways downstream of receptor activation. In addition, ligand is often difficult or expensive to obtain in useful quantities.

Probably the greatest drawback is that there are no known agonist ligands for many proteins identified as receptors. The so-called orphan receptors have usually been identified as receptors based upon protein sequence or structural similarity with known receptors. The ability to activate receptors, orphan or otherwise, without the necessity for agonist ligand will facilitate the identification of agents that modulate the intracellular events regulated by the receptor.

Constitutively active receptors couple to their downstream signaling pathway in the absence of agonist occupation of the receptor. This allows them to be used for the study of receptor signaling and function. They are also useful as components of screening platforms in drug discovery for the detection of inverse agonists or weak partial agonists, which cannot be identified using the wild type receptor. The generation of constitutively active mutants of receptor proteins is, therefore, of key importance for use of these proteins in drug discovery.

Other desired activities in addition to constitutive activation include, for example, increased or decreased binding affinity for modulators (including agonists, inverse agonists, or antagonists), which can permit the development of better modulators of a given receptor.

The use of constitutively active mutants (CAMs) in drug discovery is disclosed in US 5,750,353 which describes use of CAMs of cholecystokinin (CCK) receptors for the identification of novel compounds of either the positive agonist, the partial agonist or the inverse agonist type.

The use of constitutively active G Protein coupled receptors (GPCRs) is disclosed in US 5,882,944. Here, constitutive activation by agonist exposure is used in a method to identify inverse agonists.

The use of CAMs to study GPCRs has previously been restricted to those receptors that have already been characterized and for which agonist ligands are available. Conventional methods of generating receptor mutants, including CAMs, include random and scanning mutagenesis methods. However, these methods require the generation and screening of large numbers of mutants. Alternative methods use conformational prediction, but such methods are computationally intensive and based upon structural models whose accuracy is often uncertain. The accuracy of such prediction methods can, therefore, be questionable.

The use of pairwise alignments for the prediction of which amino acid residues to mutate to provide a protein with a desired activity or altered activity has been proposed. By examining reported studies of both naturally occurring CAMs and mutations which have been engineered into receptor proteins, certain residues can be selected which appear to be of key importance in the maintenance of receptor conformation and induction of constitutive activity. For example, the "KKXXK" motif at the C-terminal end of intracellular loop III is a well conserved motif amongst the rhodopsin-like GPCRs. Extensive studies of this region in the α_{1b}-adrenoceptor demonstrated that A293 of the KKXXK region, when substituted with every other residue, resulted in constitutive activation; the greatest level of activation were achieved using lysine and glutamate (Kjelsberg et al., 1992, *J. Biol. Chem.*, 267, 1430). Studies in other GPCRs have demonstrated that residues analogous to A293 of the α_{1b}-adrenoceptor also result in constitutive activation, including mutants of the α_{1c}-adrenoceptor (A293; McWhinney et al., *J. Biol. Chem.*, 275, 2087-2097), α_{2a}-adrenoceptor (T348; Ren et al., 1993, *J. Biol. Chem.*, 268, 16483), β2-adrenoceptor (L272; Samama et al., 1993, *J. Biol. Chem.*, 268, 4625), thyrotropin receptor (A623; Parma et al., 1993, *Nature*, 365, 649), 5-HT_{2A} (C322; Parma et al., 1998, *J. Pharmacol. Exp. Ther.*, 286, 85) and Platelet-activating receptor (L231; Parent et al., 1996, *J. Biol. Chem.*, 271, 7949). Thus, mutations within the KKXXK motif have a proven track record for inducing constitutive activity throughout a fairly diverse group of GPCRs.

It is not always the case that analogous mutations in different receptors result in constitutive activation. This is illustrated by mutations in the 'DRY' motif of the GPCRs, a region located at the junction of transmembrane domain III and intracellular loop II (see Figure 1), which is extremely well conserved across the entire GPCR super-family. This conservation appears to permit the simple prediction of residues within this region to mutate. In the α_{1b}-adrenoceptor, mutations at R143 of the DRY motif exhibit a profound effect on agonist potency, resulting in constitutive activation (Scheer et al., 1996, *EMBO J.*, 15, 3566). These results have

been reproduced in other receptors, such as rhodopsin (Cohen et al., 1993, *Biochemistry*, 32, 6111) and the LHCG receptor (Wang et al., 1993, *Mol. Endocrinol.*, 7, 85). However, the analogous mutation (D172) in the 5-HT2 receptor resulted in a reduction of agonist-induced activity and no constitutive activity (Wang et al., 1993, *Mol. Pharmacol.*, 43, 93 1). Thus, pairwise sequence comparison is an unreliable method for predicting mutations that will lead to constitutive activation or other desired activity modification.

Receptors are classified into families and sub-families based upon structural and functional characteristics of the receptor proteins. For example, the G-protein coupled receptor superfamily is subdivided into the Rhodopsin-like (also known as type 1), Secretin-like (also known as type 2), type 3 and "frizzled" type families. The Rhodopsin-like family is further sub-divided into the amine/muscarinic receptors, chemokine receptors, prostanoid receptors, and opioid/somatostatin receptors, among others. Prediction of which mutations will cause constitutive activation of orphan receptors is especially difficult because it is often not known which family of receptors an orphan receptor belongs to, or its specificity of coupling or tissue distribution.

There is a need, therefore, to provide methods for accurately predicting which amino acid residues of a protein to mutate to provide it with a desired activity or altered activity. Methods which accurately predict such mutations will obviate the need to generate large numbers of mutations which have to be screened to in order to identify mutants showing the desired activity.

In its broadest sense, the invention provides methods for predicting which mutations to make to a receptor protein to provide it with a desired activity or altered activity. According to the invention, phylogenetic profiling techniques are used to assign the receptor protein to a family or subfamily. The sequences of known members of the family or subfamily are aligned with the sequence of the protein to generate a multiple sequence alignment. This process facilitates accurate assignment of residues in the protein which are analogous to residues in known members of the family or subfamily that were previously mutated. Knowledge of the effects of mutations previously made to or identified in the known members of the family or subfamily is used to predict which mutations to make to the protein. That is, the residue or residues in the protein which are selected for mutation are those which are analogous to residues which have previously been mutated or identified in mutants in the known family or subfamily members, and which altered the activity of these known family or subfamily members. Usually, residues in the protein that are analogous to residues of the known family or subfamily members

which, when mutated, caused different effects in different members, or which did not have any effect in some members, will not be selected for mutation.

SUMMARY OF THE INVENTION

The invention encompasses a method of predicting a site for mutation of a first cellular receptor wherein the mutation alters the activity of the first cellular receptor, comprising the steps of: (a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same receptor family; (b) allocating the first cellular receptor to a receptor sub-family; and (c) selecting an amino acid residue of the first cellular receptor for mutation, wherein the amino acid residue is analogous to a residue, the mutation of which is known to cause altered activity of a second cellular receptor, whereby a site for mutation of the first cellular receptor is predicted.

In one embodiment, the first and second cellular receptors are cell surface receptors.

In another embodiment, the first and second cellular receptors are G-protein coupled receptors.

In a preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another preferred embodiment, the step of allocating the first cellular receptor to a receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of a human α_{1B} adrenoceptor.

In another preferred embodiment, the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

In another preferred embodiment, the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as the first G protein-coupled receptor.

In another preferred embodiment, the second G protein-coupled receptor is in the same sub-family as the first G protein-coupled receptor.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

The invention further encompasses a method of obtaining a mutant of a first cellular receptor, wherein the mutant has altered activity as compared to a wild type cellular receptor, comprising the steps of: (a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same family; (b) allocating the first cellular receptor to a cellular receptor sub-family; (c) selecting an amino acid residue of the first cellular receptor for mutation, wherein the selected amino acid residue is analogous to a residue, the mutation of which is known to cause altered activity of a second cellular receptor; (d) mutating the selected amino acid residue of the cellular receptor; and (e) expressing the mutated cellular receptor in a cell.

In one embodiment, the first and second cellular receptors are cell surface receptors.

In another embodiment, the first and second cellular receptors are G-protein coupled receptors.

In a preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another embodiment, the step of allocating the first cellular receptor to a cellular receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

In another preferred embodiment, the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

In another preferred embodiment, the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as the first cellular receptor.

In another preferred embodiment, the second cellular receptor is in the same sub-family as the first cellular receptor.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

The invention further encompasses a method of predicting a site for mutation of a cellular receptor wherein the mutation alters the activity of the cellular receptor, the method comprising the steps of: (a) performing a multiple sequence alignment of the cellular receptor with other cellular receptors in the same receptor family; (b) allocating the cellular receptor to a cellular receptor sub-family; (c) determining whether mutant data are available for a member of the sub-family of the cellular receptor; wherein if mutant data are available, then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue, the mutation of which is known to cause altered activity of a member of the same sub-family as the cellular receptor, whereby a site for mutation of the cellular receptor is predicted.

In one embodiment, the first and second cellular receptors are cell surface receptors.

In another embodiment, the first and second cellular receptors are G-protein coupled receptors.

In a preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another preferred embodiment, the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

In another preferred embodiment, the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

In another preferred embodiment, the step allocating the cellular receptor to a cellular receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the sub-family of the first cellular receptor.

In another preferred embodiment, the second cellular receptor is in the same sub-family as the first cellular receptor.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

The invention further encompasses a method of predicting a site for mutation of a G protein-coupled receptor wherein the mutation alters the activity of the G protein-coupled receptor, comprising the steps of: (a) performing a multiple sequence alignment of the G protein-coupled receptor with other G protein-coupled receptors; (b) allocating the G protein-coupled receptor to a G protein-coupled receptor sub-family; (c) determining whether mutant data are available for a member of the same sub-family as the G protein-coupled receptor; wherein if mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a member of the same sub-family as the G protein-coupled receptor; and wherein if mutant data are not available then selecting an amino acid residue for mutation by identifying an amino acid residue selected from the group consisting of a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor and an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor; andwhereby a site for mutation of a G protein-coupled receptor is predicted.

In a preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another preferred embodiment, the step of allocating the G protein-coupled receptor to a G protein-coupled receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

The invention further encompasses a method of obtaining a mutant of a cellular receptor, wherein the mutant has altered activity as compared to a wild type cellular receptor, comprising the steps of: (a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same family; (b) allocating the cellular receptor to a cellular receptor sub-family; (c) determining whether mutant data are available for a member of the same sub-family as the cellular receptor; wherein when mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a member of the same sub-family as the

cellular receptor; (d) mutating the selected amino acid residue of the cellular receptor; and (e) expressing the mutated cellular receptor in a cell.

In a preferred embodiment, the first and second cellular receptors are cell surface receptors.

In another preferred embodiment, the first and second cellular receptors are G-protein coupled receptors.

In another preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another preferred embodiment, the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

In another preferred embodiment, the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

In another preferred embodiment, the step of allocating the cellular receptor to a cellular receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as the first cellular receptor.

In another preferred embodiment, the second cellular receptor is in the same sub-family as the first cellular receptor.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

The invention further encompasses a method of obtaining a mutant of a G protein-coupled receptor, wherein the mutant has altered activity as compared to a wild type G protein-coupled receptor, comprising the steps of: (a) performing a multiple sequence alignment of the

first G protein-coupled receptor with other G protein-coupled receptors; (b) allocating the G protein-coupled receptor to a G protein-coupled receptor sub-family; (c) determining whether mutant data are available for a member of the same sub-family as the G protein-coupled receptor; wherein if mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a member of the same sub-family as the G protein-coupled receptor; and wherein if mutant data are not available then selecting an amino acid residue for mutation by identifying an amino acid residue selected from the group consisting of a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor and an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human $_{1B}$ adrenoceptor; (d) mutating the selected amino acid residue of the G protein-coupled receptor; and (e) expressing the mutated G protein-coupled receptor in a cell.

In a preferred embodiment, the first G protein-coupled receptor is an orphan receptor.

In another preferred embodiment, the step of allocating the G protein-coupled receptor to a G protein-coupled receptor sub-family is performed by phylogenetic profiling.

In another preferred embodiment, the mutation causes the receptor to become constitutively activated.

In another preferred embodiment, the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

In another preferred embodiment, the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

In another preferred embodiment, the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as the first cellular receptor.

In another preferred embodiment, the second cellular receptor is in the same sub-family as the first cellular receptor.

The invention further encompasses a mutated GPR8 receptor comprising altered activity as compared to a wild type GPR8 receptor, wherein the GPR8 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 124 from aspartate to alanine, a mutation at amino acid 127 from asparagine to alanine and a mutation at amino acid 259 from threonine to glutamate.

The invention further encompasses a mutated GPR7 receptor comprising altered activity as compared to a wild type GPR7 receptor, wherein the GPR7 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 116 from aspartate to alanine, a mutation at amino acid 119 from asparagine to alanine and a mutation at amino acid 250 from threonine to glutamate.

The invention further encompasses a mutated GPR10 receptor comprising altered activity as compared to a wild type GPR10 receptor, wherein the GPR10 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 244 from tyrosine to glutamate and a mutation at amino acid 247 from valine to glutamate.

The invention further encompasses a mutated GPR1 receptor comprising altered activity as compared to a wild type GPR1 receptor, wherein the GPR1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 176 from arginine to alanine, a mutation at amino acid 245 from phenylalanine to glutamate, and a mutation at amino acid 120 from asparagine to alanine.

The invention further encompasses a mutated GPR17 receptor comprising altered activity as compared to a wild type GPR17 receptor, wherein the GPR17 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 114 from asparagine to alanine and a mutation at amino acid 234 from valine to glutamate.

The invention further encompasses a mutated GPR4 receptor comprising altered activity as compared to a wild type GPR4 receptor, wherein the GPR4 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 100 from asparagine to alanine and a mutation at amino acid 223 from lysine to glutamate.

The invention further encompasses a mutated GPR15 receptor comprising altered activity as compared to a wild type GPR15 receptor, wherein the GPR15 receptor comprises a mutation

selected from the group consisting of a mutation at amino acid 116 from asparagine to alanine and a mutation at amino acid 240 from isoleucine to glutamate.

The invention further encompasses a mutated GPR20 receptor comprising altered activity as compared to a wild type GPR20 receptor, wherein the GPR20 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 133 from asparagine to alanine and a mutation at amino acid 240 from methionine to glutamate.

The invention further encompasses a mutated HM74 receptor comprising altered activity as compared to a wild type HM74 receptor, wherein the HM74 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 110 from asparagine to alanine and a mutation at amino acid 230 from isoleucine to glutamate.

The invention further encompasses a mutated OGR1 receptor comprising altered activity as compared to a wild type OGR1 receptor, wherein the OGR1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 104 from asparagine to alanine and a mutation at amino acid 227 from glutamine to glutamate.

The invention further encompasses a mutated EBI2 receptor comprising altered activity as compared to a wild type EBI2 receptor, wherein the EBI2 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 114 from asparagine to alanine and a mutation at amino acid 243 from leucine to glutamate.

The invention further encompasses a mutated BONZO receptor comprising altered activity as compared to a wild type BONZO receptor, wherein the BONZO receptor comprises a mutation selected from the group consisting of a mutation at amino acid 112 from asparagine to alanine and a mutation at amino acid 230 from leucine to glutamate.

The invention further encompasses a mutated RDC1 receptor comprising altered activity as compared to a wild type RDC1 receptor, wherein the RDC1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 127 from asparagine to alanine and a mutation at amino acid 251 from arginine to glutamate.

The invention further encompasses a mutated O15218 receptor comprising altered activity as compared to a wild type O15218 receptor, wherein the O15218 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 136 from asparagine to alanine and a mutation at amino acid 257 from cysteine to glutamate.

The invention further encompasses a mutated H963 receptor comprising altered activity as compared to a wild type H963 receptor, wherein the H963 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 97 from asparagine to alanine and a mutation at amino acid 222 from leucine to glutamate.

The invention further encompasses a mutated GPR30 receptor comprising altered activity as compared to a wild type GPR30 receptor, wherein the GPR30 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 140 from asparagine to alanine and a mutation at amino acid 258 from leucine to glutamate.

The invention further encompasses a mutated GPR2 receptor comprising altered activity as compared to a wild type GPR2 receptor, wherein the GPR2 receptor comprises a mutation at amino acid 238 from leucine to glutamate.

The invention further encompasses a mutated GPR5 receptor comprising altered activity as compared to a wild type GPR5 receptor, wherein the GPR5 receptor comprises a mutation at amino acid 224 from valine to glutamate.

The invention further encompasses a mutated GPR13 receptor comprising altered activity as compared to a wild type GPR13 receptor, wherein the GPR13 receptor comprises a mutation at amino acid 230 from isoleucine to glutamate.

The invention further encompasses a mutated GPR18 receptor comprising altered activity as compared to a wild type GPR18 receptor, wherein the GPR18 receptor comprises a mutation at amino acid 231 from isoleucine to glutamate.

The invention further encompasses a mutated GPR21 receptor comprising altered activity as compared to a wild type GPR21 receptor, wherein the GPR21 receptor comprises a mutation at amino acid 251 from alanine to glutamate.

The invention further encompasses a mutated GPR22 receptor comprising altered activity as compared to a wild type GPR22 receptor, wherein the GPR22 receptor comprises a mutation at amino acid 312 from phenylalanine to glutamate.

The invention further encompasses a mutated GPR25 receptor comprising altered activity as compared to a wild type GPR25 receptor, wherein the GPR25 receptor comprises a mutation at amino acid 242 from leucine to glutamate.

The invention further encompasses a mutated GPR31 receptor comprising altered activity as compared to a wild type GPR31 receptor, wherein the GPR31 receptor comprises a mutation at amino acid 221 from glutamine to glutamate.

The invention further encompasses a mutated GPR38 receptor comprising altered activity as compared to a wild type GPR38 receptor, wherein the GPR38 receptor comprises a mutation at amino acid 297 from valine to glutamate.

The invention further encompasses a mutated GPR39 receptor comprising altered activity as compared to a wild type GPR39 receptor, wherein the GPR39 receptor comprises a mutation at amino acid 282 from isoleucine to glutamate.

The invention further encompasses a mutated GPR40 receptor comprising altered activity as compared to a wild type GPR40 receptor, wherein the GPR40 receptor comprises a mutation at amino acid 223 from alanine to glutamate.

The invention further encompasses a mutated GPR41 receptor comprising altered activity as compared to a wild type GPR41 receptor, wherein the GPR41 receptor comprises a mutation at amino acid 224 from alanine to glutamate.

The invention further encompasses a mutated GPR42 receptor comprising altered activity as compared to a wild type GPR42 receptor, wherein the GPR42 receptor comprises a mutation at amino acid 224 from alanine to glutamate.

The invention further encompasses a mutated GPR43 receptor comprising altered activity as compared to a wild type GPR43 receptor, wherein the GPR43 receptor comprises a mutation at amino acid 221 from valine to glutamate.

The invention further encompasses a mutated MGR receptor comprising altered activity as compared to a wild type MGR receptor, wherein the MGR receptor comprises a mutation at amino acid 263 from tyrosine to glutamate.

The invention further encompasses a method of identifying a compound which modulates the activity of a receptor mutated according to a method of the invention comprising a) contacting a candidate compound with the receptor, and b) determining activity of the receptor in the presence of the compound, wherein a difference in receptor activity in the presence and absence of the candidate compound is indicative of compound modulation.

In a preferred embodiment, the compound is further determined to be an inverse agonist, partial agonist or an agonist of the receptor activity.

As used herein, the term "predicting a site for mutation" means identifying an amino acid or group of amino acids in a given polypeptide that, if mutated, are likely to change the function of that polypeptide in a desired manner.

As used herein, the term "cellular receptor" means a polypeptide that interacts with a extracellular signaling factor and thereby mediates the cellular response to that signaling factor. Cellular receptors include "cell surface receptors", in which at least a portion of the receptor is exposed to the extracellular surroundings. In addition, cell surface receptors include those receptors, such as the steroid hormone receptor, that interact with intracellularly with signaling molecules that pass into the cell without necessarily interacting with a cell surface receptor.

As used herein, the term "mutation" refers to a change in the amino acid sequence of a polypeptide relative to the sequence of that polypeptide occurring most frequently in nature. A mutation may be a deletion, substitution or insertion of one or more amino acids relative to the wild-type amino acid sequence.

As used herein, the term "wild-type" refers to the form of a protein or receptor one wishes to mutate, as it exists prior to that mutation.

As used herein, the "activity" of a receptor refers to the function of the receptor in mediating a cellular response to an extracellular signal. Non-limiting examples of functions that constitute activity include stimulation of GDP for GTP exchange in a G-protein, kinase activation, protease activation, phosphatase activation and stimulation of protein:protein interaction. In addition, the "activity" of a receptor is reflected in the signaling function or the activity of downstream signaling pathways, or ultimately, in changes in the expression of one or more genes. Wild-type receptor activity is regulated, in that the activity is modulated in response to a ligand or agonist. A mutant receptor exhibits "constitutive" activity if it is active in the absence of a ligand or agonist. Constitutive activity need not be particularly high level, but must be greater than the activity level of the wild type receptor in the absence of ligand or agonist.

The "activity" of a non-receptor protein being targeted for mutation according to the invention refers to a measurable activity of that protein that is related to the function of that protein. For example, for an enzyme, activities include binding of substrate or substrates, binding of or interaction with regulatory factors or subunits, actual conversion of substrate to product, or a downstream effect that is proportional to the enzymatic activity (e.g., a change in gene expression or accumulation or decrease in the concentration of a metabolite).

Mutations predicted by the methods of the invention result in "altered" activity, which alteration may be an increase or decrease relative to the activity of the protein before that mutation, but preferably results in generation of constitutive activity. The activity of a receptor or other protein is "increased" as used herein if at least one measure of its activity (e.g., kinase activity, downstream signaling, transcription activity, etc.) is at least 10% greater and preferably 20%, 30%, 50%, 75% or greater, or even 100% (2-fold), 3-fold, 5-fold, 10-fold greater or more, relative to a wild type or unstimulated receptor. Conversely, the activity of a receptor or other protein is "decreased" as used herein if at least one measure of its activity is at least 10% less and preferably 20%, 30%, 50%, 75%, 85%, 90%, 95%, 98% or even up to and including 100% (no activity) less than the activity of either 1) a wild type receptor in response to a given stimulus, or 2) the activity of a constitutively active mutant of a receptor.

As used herein, the term "allocating a cellular receptor to a sub family" refers to the process whereby the closest relatives of a given cellular receptor are identified from within the larger group of relatives or family members.

As used herein, the term "G-Protein coupled receptor," or "GPCR" refers to a membrane associated polypeptide with 7 alpha helical transmembrane domains. Functional GPCR's associate with a ligand or agonist and also associate with and activate G-proteins. However, placement of a polypeptide the "GPCR family" does not required demonstration of ligand or agonist binding or association with G-proteins.

As used herein, the term "transmembrane domain" refers to a hydrophobic stretch of amino acids in a polypeptide which is inserted into the cell membrane. A transmembrane domain can form the junction between intracellular and extracellular domains of a polypeptide, or it can serve as the membrane-anchoring sequence of a polypeptide that is only exposed to one or the other side of the membrane. Transmembrane domains are predicted in proteins according to hydropathy plots as described by, for example, Kyle and Doolittle (1982, J. Mol. Biol. 157: 105-32) or by computer algorithms incorporating the same principles (e.g., Peptide_Structure).

As used herein, the term "orphan receptor" refers to a polypeptide that is identified as a receptor on the basis of structural characteristics (e.g., seven alpha-helical transmembrane domains), but for which a ligand or agonist has not been identified.

As used herein, the term "analogous" when used in reference to an amino acid residue refers to the position of an amino acid in a given sequence relative to the position of a similar or identical amino acid in the sequence of a close relative. Analogous positions are identified by multiple sequence alignment.

As used herein, the term "conserved" when used is reference to an amino acid in a polypeptide sequence means that an identical or similar amino acid is located at an analogous position in one or more other members of a family.

As used herein, the term "determining whether mutant data are available for a member of the same sub-family" refers to the process of identifying published data describing one or more mutants of a given protein identified as a member of the same sub-family as a protein of interest. The existence of available data may be determined by one of skill in the act by searching any of a number of databases, including, but not limited to Medline, US or Foreign Patent databases, online databases such as OMIM (Online Mendelian Inheritance in Man) or subscription database and search providers, such as those available through DIALOG®. Keyword searches with the names of identified close relatives in these or another databases known in the art are expected to identify known mutants of those proteins. The data identified in this manner should teach the desired alteration in the activity of the relative (e.g., constitutive activation).

As used herein, the term "agonist" means a molecule or composition that binds to and increases the activity of a receptor. An agonist includes, but is not limited to the natural ligand for a receptor. When an agonist is not the actual ligand, it can increase receptor activity in the absence of the natural ligand.

As used herein, the term "inverse agonist" means a molecule or composition that decreases the activity of a receptor below the baseline activity that receptor has in the absence of ligand or agonist. Unlike an antagonist, an inverse agonist does not function by blocking the activation by an agonist. Rather, an inverse agonist, on its own, reduces the baseline activity of the receptor.

As used herein, the term, "partial agonist" means a molecule or composition that increases the activity of a receptor relative to the receptor's activity in the absence of the molecule or composition, but where the increase is less than that induced by an agonist at similar levels of receptor binding. One skilled in the art can determine whether a given compound is an agonist, an inverse agonist or a partial agonist of a given receptor.

CAM orphan GPCRs identified using methods of the invention may be used in methods for determining the function of the orphan GPCR as a consequence of its ability to constitutively couple to downstream signaling pathways when it is expressed in native cells. CAM orphan GPCRs identified using methods of the invention may also be used in screening assays for the detection of inverse agonists and weak partial agonists, which would otherwise not be detectable using the wild type receptor.

It will be appreciated that the invention is not limited to methods of predicting mutations of orphan GPCRs which will constitutively activate these GPCRs. Methods of the invention can be applied to other GPCRs and to other families of receptor proteins. According to the invention, the use of phylogenetic analysis to inform decision making on mutagenesis may be used for any family of receptor proteins. An example of another receptor protein family to which the methods of the invention could be applied is the nuclear receptor family of transcription factors such as the estrogen receptor family. Mutation of an estrogen receptor has rendered that receptor capable of signaling in the absence of hormone stimulation (White at al., 1997, *EMBO J., 16,* 1427).

The method of the invention is useful for the prediction of mutations in any family of proteins, receptor or otherwise. It will also be appreciated that methods of the invention are not limited to prediction of mutations which cause constitutive activation of a protein. Mutations which cause other desired activities or altered activities may be predicted by methods of the invention. Other desired activities include increases or decreases in agonist, inverse agonist or antagonist binding affinity, agonist or inverse agonist potency, cell surface expression, receptor stability, basal second messenger stimulation, and maximal agonist stimulated second messenger level.

Methods of the invention are thought to provide more accurate predictions of which mutations to make to a protein to provide it with a desired activity or altered activity than known methods. Methods of the invention also prevent the need to generate and screen large numbers of mutants in order to look for mutants having a desired activity or altered activity.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a diagram of a G-protein coupled receptor. TM1 – TM7 are the transmembrane domains, E1 – E3 are the extracellular loop domains, and I1 – I4 are the intracellular domains.

Figure 2 shows a multiple sequence alignment of selected human GPCRs. The four-letter code for protein and receptor names used in the Swiss-Prot database is used to refer to the proteins in the figure (see Table 1). a1ab is α_{1b} adrenoceptor; ag2r is angiotensin 2 receptor 1, gpr8 is the GPR8 orphan receptor; nk1r is the neurokinin 1 receptor; gpr1 is the GPR1 orphan receptor; c5ar is the C5 anaphylotoxin receptor; gasr is the gastrin receptor; gpra is the GPR10

orphan receptor. Residues marked * were aligned manually and intervening portions of the alignment constructed automatically using PILEUP (GCG). *1* and *2* indicate residues described in the text where mutagenesis has led to constitutive activity in several diverse receptors.

Figure 3a-3d shows a phylogenetic tree of the rhodopsin family of GPCRs. Figure 3a shows the whole phylogenetic tree, and figures 3b-3d shows enlargements of the regions of the tree containing the amine/muscarinic receptors, prostanoid receptors, and chemokine receptors, respectively. The tree was constructed by an evolutionary distance method using PROTIST and FITCH (Phylip). Orphan receptors are in the smallest circles.

Figures 4-8 show alignments of orphan GPCRs with GPCRs in which activating mutants are known. Sites of activating mutants are indicated in ***bold italics***. Proposed substitutions at analogous sites are shown below the relevant sequence. Conserved residues used to anchor multiple sequence alignments are indicated by an asterisk (*).

Figure 4 shows a multiple sequence alignment of a portion of the GPR8 orphan receptor sequence from amino acids 113 to 153 with homologous portions of sequence from other members of the opioid receptor family. The four-letter code for protein and receptor names used in the Swiss-Prot database is used to refer to the proteins in the figure (see Table 1). Oprd, m and k are, respectively, the delta, mu and kappa-opioid receptors. Oprx is the nociceptin receptor . Ssr1-5 refer respectively to the somatostatin receptors types 1-5.

Figure 5 shows an alignment of: A) a portion of the GPR8 orphan receptor from amino acids 255 to 280 with amino acids 289 to 314 of the alpha-1b-adrenoceptor - the residue analogous to A293 is identifiable according to method 1; and B) a portion of the GPR8 orphan receptor from amino acids 117 to 143 with amino acids 101 to 127 of the type 1A angiotensin II receptor. The residue analogous to N111 is identifiable according to method 2.

Figure 6 shows an alignment of: A) a portion of the GPR10 orphan receptor from amino :acids 272 to 297 with amino acids 289 to 314 of the alpha-1b-adrenoceptor - from which the residue analogous to A293 is identifiable according to method 1; and B) a portion of the GPR10 orphan receptor from amino acids 236 to 251 with amino acids 208 to 223 of the neurokinin NK1 receptor, in which the residue Y216, when mutated to glutamate, gives rise to increased agonist affinity.

Figure 7 shows an alignment of: A) a portion of the GPR1 orphan receptor from amino acids 234 to 266 with amino acids 289 to 314 of the alpha-1b-adrenoceptor - from which the residue analogous to A293 is identifiable according to method 1; B) a portion of the GPR1 orphan receptor from amino acids 110 to 136 with amino acids 101 to 127 of the type 1A angiotensin II receptor - from which the residue analogous to N111 is identifiable according to method 2; and C) a portion of a multiple sequence alignment of GPR1, showing amino acids 162-187 (SEQ ID NO:N) with chemoattractant receptors c5 anaphylotoxin receptor (amino acids 170-188, SEQ ID NO:N), fMET peptide receptor 1 (amino acids 150-176, SEQ ID NO:N) and the leukotriene receptor designated P2Y7 (amino acids 142-168, SEQ ID NO:N).

Figure 8 shows an alignment of: A) a portion of the GPR17 orphan receptor from amino acids 230 to 255 with amino acids 289 to 314 of the alpha-1b-adrenoceptor - from which the residue analogous to A293 is identifiable according to method 1; and B) a portion of the GPR17 orphan receptor from amino acids 104 to 130 with amino acids 101 to 127 of the type 1A angiotensin II receptor, from which the residue analogous to N111 is identifiable according to method 1.

Figure 9 shows an alignment of a portion of the GPR21 orphan receptor from amino acids 247 to 272 with amino acids 289 to 314 of the alpha-1b-adrenoceptor, from which the residue analogous to A293 is identifiable according to method 1. Residues marked * were aligned manually. Residues in bold italics are the sites of known mutations leading to constitutive activity. The predicted substitution for GPR21 is shown below the sequence..

Figure 10 shows a flow diagram illustrating the procedure followed in the mutation prediction method of the invention. Computational processes are in square boxes, queries in oval boxes and laboratory procedures in hexagonal boxes.

Table I provides a list of GPCRs, abbreviations for the receptors, the Swiss-Prot I.D. and GenBank Accession Numbers for proteins referred to in the table.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides methods of predicting the sites to mutate in order to alter the activity of a cellular receptor in a desired manner. The method is suitable for the prediction of sites to mutate in any type of receptor that may be placed into a family of related receptors.

Receptors, methods of placing the receptors into a family and/or sub-family based on sequence alignments and phylogenetic profiling, methods of predicting sites to mutate on the basis of these steps, methods of making the desired mutations and methods of determining whether a mutation successfully alters receptor activity are described herein below.

A. Receptors to be Mutated Using the Methods of the Invention

Any cellular receptor, including cell surface receptors and intracellular receptors may be targeted for mutation through the methods of the invention. Receptors are categorized according to their structures, as well as by the nature of the second messenger pathways through which they provide signaling. The major types of receptors include G-protein coupled receptors, receptors with intrinsic enzymatic activity (e.g., tyrosine kinase, ser/thr kinase or phosphatase), ligand-gated ion channels, and immune recognition receptors. The characteristics of members of each of these groups, including downstream signalling targets, are described below.

1. G-Protein Coupled Receptors

G-protein-coupled receptors are distinguished by the presence of seven alpha helical transmembrane domains and the interaction of an intracellular domain of the proteins with GTP-binding proteins, or G-proteins. One database (the G-Protein Coupled Receptor Database, maintained by the University of Texas Health Sciences Center), had over 2400 entries, ranging over the evolutionary spectrum from bacteria to mammals, as of mid-1999. Proteins are placed into the GPCR superfamily due to the presence of the seven predicted alpha helical transmembrane domains, and may not necessarily have any appreciable similarity at the amino acid sequence level. The accepted nomenclature for the domains refers to the transmembrane domains as TM1 - TM7, the intracellular loops as I1 – I4, and the extracellular loops as E1 –E3, with the domain preceding TM1 being the N terminus and the domain following I4 being the C terminus. Transmembrane domains are identified in a polypeptide sequence using any of a number of computer algorithms, e.g. PEPTIDESTRUCTURE [Genetics Computer Group (GCG), Madison, WI], that predict secondary structure and hydropathy. As the name implies, many known GPCRs interact with and activate G-proteins. However, the actual interaction of the receptor with a G-protein need not have been established in order to place a receptor into the superfamily.

GPCRs have been classified into 5 major groups, Classes A-E. Class A receptors are related to Rhodopsin and the β2-adrenergic receptors. Class A is sub-divided into groups I-VI,

as follows. Group I includes the olfactory receptors, adenosine receptors, melanocortin receptors, and others. Group II includes the biogenic amine receptors (muscarinic, adrenergic, serotonin, etc). Group III includes the vertebrate opsins and neuropeptide receptors. Group IV includes the invertebrate opsins. Group V includes the receptors for chemokines, chemotactic signals, somatostatin, opioids and others . Group VI includes the melatonin receptors.

Class B contains receptors most closely related to the calcitonin, parathyroid hormone/parathyroid hormone related peptide and glucagon receptors. Class B is sub-divided into Groups I-IV as follows. Group I includes the calcitonin, calcitonin-like, corticotropin releasing factor and diuretic hormone receptors, as well as a number of orphan receptors. Group II includes the parathyroid hormone/parathyroid hormone related peptide receptors and related orphan receptors. Group III includes the glucagon-like peptide receptors, glucagon receptors, gastric inhibitory peptide receptor subtypes, pituitary adenylyl cyclase activating peptide receptors, vasoactive intestinal peptide receptors, secretin receptors, and a number of orphan receptors. Group IV includes the latrotoxin receptors and a number of orphan receptors.

Class C receptors include the extracellular calcium-sensing receptors and the metabotropic glutamate receptors. Class C is sub-divided into Groups I-IV, as follows. Group I includes the metabotropic glutamate receptors. Group II includes the calcium receptors. Group III includes the GABA-B receptors. Group IV includes the putative pheromone receptors.

Class D receptors include those related to the yeast mating type STE2 pheromone receptor.

Class E receptors include those related to the yeast mating type STE3 pheromone receptor.

It is possible, likely even, that further sub-families of GPCRs will be identified, or that as data regarding natural ligands and transduction pathways emerges, existing families and sub-families will be re-arranged. Because it is based upon phylogenetic relationships, the method of the invention is applicable to all members of the superfamily, both existing and yet to be discovered.

GPCRs are able, in response to agonist binding, to stimulate the exchange of bound GDP for GTP on the α-subunit of associated G-proteins. The GTP-bound form of the G-protein is the active form, with the capacity to activate the downstream signaling cascade that ultimately

mediates the cell's response to ligand or agonist binding of the GPCR. It is thought that the intracellular loop joining transmembrane domains 5 and 6 of the receptor interacts with the G-protein. G-protein activation may be monitored as a measure of receptor activation using, for example, assays to detect GTP binding to the membrane comprising the receptors or by direct measurement of GTPase activity.

Downstream receptor signaling pathways used by GPCRs include modulation of adenylate cyclase activity, which alters the intracellular levels of the second messenger cyclic AMP (cAMP). The modulation of intracellular cAMP levels may occur in several different ways. For example, activation of adenylate cyclase, leading to an increase in cAMP, occurs upon the activation of β-adrenergic, glucagon and odorant molecule receptors. Other receptors are coupled not to activating G-proteins, but to inhibitory G-proteins that repress adenylate cyclase activity upon receptor activation. These receptors include the α-type adrenergic receptors. The downstream effects of cAMP production also differ for the different receptors that modulate cAMP levels. For example, activation of adenylate cyclase by the β-adrenergic and glucagon receptors results in Protein Kinase A (PKA) activation, whereas the activation of adenylate cyclase by the odorant receptors induces activation of ion channels.

The activity of GPCRs that modulate adenylate cyclase activity may therefore be measured by the measurement of adenylate cyclase enzymatic activity or by the measurement of cAMP levels. Other means of monitoring the activation of GPCRs that function through adenylate cyclase modulation include measurement of cAMP-responsive gene expression using reporter assays (see below).

Another downstream receptor signaling mechanism used by members of the GPCR superfamily is the activation of Phospholipase C-γ (PLC-γ), which causes the hydrolysis of polyphosphoinositides, which generates the second messengers diacylglycerol (DAG) and inositol triphosphate, which subsequently regulate the activity of further participants in the signal transduction cascade. Receptors using this pathway include, for example, the bradykinin, vasopressin and angiotensin receptors. Measurement of phosphoinositol breakdown, DAG and inositol triphosphate levels may be used to monitor activation of PLC-γ-activating GPCRs. In addition, the production of DAG and the release of intracellular calcium stores as a result of PLC-γ activation can activate Protein Kinase C (PKC). The activity of PKC may thus be measured in order to monitor GPCR activation of PLC-γ. Alternatively, reporter gene assays sensitive to the levels of PLC-γ second messengers or to PKC activation may be used.

Another sub-family of GPCRs, the photoreceptors, is coupled to the transducin G-protein. Activation of transducin leads to the activation of a phosphodiesterase that reduces cyclic GMP (cGMP) levels, which results in the closing of an ion channel, resulting in the hyperpolarization of the cell. Hyperpolarization of the cell in response to receptor activation may be monitored by, for example, patch clamp techniques.

Each of the signal transduction pathways used by the GPCRs has multiple downstream elements, any one of which may be inappropriately regulated in a disease state. The constitutive activation of the receptors using mutations predicted according to the method of the invention will permit the screening of drugs to alter such inappropriate regulation. Diseases involving GPCRs (and the receptors linked to them), include but are not limited to the following: Retinitis pigmentosa (rhodopsin);Stationary Night Blindness (rhodopsin); Color blindness (Red and Green opsins); Nephrogenic DI (V2 receptor); Isolated glucocorticoid deficiency (ACTH receptor); Hyperfunctioning thyroid adenomas (TSH-receptor); Familial precocious puberty (LH receptor); Familial hypocalciuric hypercalcemia (Ca^{+2}-sensing receptor); and Neonatal severe hyperparathyroidism (Ca^{+2}-sensing receptor).

2. Receptors with Enzyme Activity

A number of transmembrane and intracellular receptors of different families have intrinsic enzymatic activity that is directly involved in the transduction of signals. The most common, or at least the best characterized, is the group of receptors that have tyrosine kinase activity. Others have tyrosine phosphatase activity, guanylate cyclase activity or serine/threonine kinase activity. The methods of the invention are suitable for predicting mutations that will constitutively activate or otherwise alter the activity of receptors in each of these categories.

a) Tyrosine kinases

Receptor tyrosine kinases (RTKs) catalyze the phosphorylation of proteins on tyrosine residues. RTKs are characterized by the presence of four major domains: 1) an extracellular ligand binding domain; 2) an intracellular tyrosine kinase domain; 3) an intracellular regulatory domain; and 4) a transmembrane domain. The RTKs are placed into families based upon structural features in their extracellular domains including cysteine-rich domains, immunoglobulin-like domains, leucine-rich domains, kringle domains, cadherin domains, fibronectin type III domains, EGF-like domains, and the presence or absence of a non-kinase

insert in the intracellular kinase domain. The various patterns of the presence and or arrangement of these domains has resulted in the placement of the RTKs into at least 7 families or classes.

Class I RTKs are characterized by cysteine-rich sequences in the extracellular domain, and include, for example, EGF receptor, NEU/HER2 and HER3.

Class II RTKs are characterized by cysteine-rich sequences in the extracellular domain, and exist as disulfide linked heterotetramers. Class II receptors include, for example, the insulin receptor and IGF-1 receptor.

Class III RTKs are characterized by the presence of five immunoglobulin-like sequences in the extracellular domain and an insert in the kinase domain. Class III receptors include, for example, the PDGF receptors and c-Kit.

Class IV RTKs are characterized by the presence of three immunoglobulin-like sequences in the extracellular domain and the presence of a kinase insert. Class IV receptors include, for example, the FGF receptors.

Class V RTKs are characterized by the presence of seven immunoglobulin-like sequences in the extracellular domain and the presence of a kinase insert. Class V receptors include, for example, the vascular endothelial cell growth factor (VEGF) receptor.

Class VI RTKs are characterized by a disulfide-linked heterodimeric structure with cysteine-rich extracellular domains, wherein one of the two subunits is completely extracellular. Class VI receptors include hepatocyte growth factor (HGF) and scatter factor (SC) receptors.

Class VII RTKs contain no or few cysteine-rich domains. Examples include the neurotrophin receptors (trkA, trkB, trkC) and the NGF receptor. The NGF receptor has a leucine-rich domain.

A comprehensive listing of identified kinases, including receptors with kinase activity, is maintained online and updated monthly at http://www.sdsc.edu/kinases/pkr/pk_catalytic/pk_cat_list.html. Kinases on this list were identified using a panel of kinase domain query sequences, one from each of 57 tyrosine kinase sub-families defined by Hanks et al., to perform BLAST searches. A listing of diseases associated with kinases, including not only tyrosine

kinases, but also serine/threonine kinases, serine kinases and threonine kinases, is also maintained online at http://www.sdsc.edu/kinases.

Downstream signaling by many activated RTKs involves the phosphorylation of tyrosine residues on Src-homology-2-domain (SH2 domain)-containing proteins. SH2 domains are regulated protein:protein interaction domains. The phosphorylation and resulting regulated protein:protein interaction regulates the activity of the SH2-domain proteins, many of which have intrinsic enzymatic activity, including, for example, PLC-γ, c-Ras-asssociated GTPase activating protein (RasGAP), phosphatidylinositol-3-kinase (PI-3K) and protein phosphatase 1C (PTP1C), as well as intracellular tyrosine kinases, such as the Src family of tyrosine kinases. The activity of RTKs that activate these pathways may be monitored by measuring the activity of the downstream signal transduction proteins by direct measurement of enzyme activity, using, for example, labeled subtrates, or by measuring the accumulation of enzymatic reaction products, or by measuring the expression of reporter genes linked to transcriptional control sequences activated by the signal transduction cascade from that receptor.

The EGF, PDGF, NGF and insulin receptor tyrosine kinases, to name a few, are also known to activate the intracellular mitogen activated protein (MAP) kinases, as are receptors that signal through PKC activation. MAP kinase activity may be measured directly by, for example, incorporation of ^{32}P-labeled phosphate into test substrates, such as myelin basic protein. MAP kinases regulate the activity of a number of transcription factors, including serum-response factor (SRF), c-Myc, c-Fos and c-Jun, and members of the steroid/thyroid hormone receptor superfamily of proteins. The activity of receptors that signal through MAP kinases may thus be measured by measurement of the expression of reporter constructs driven by promoter sequences responsive to these transcription factors.

Mutation sites for the modification of RTKs may be predicted using the methods of the invention.

b. Receptor Tyrosine Phosphatases (RTPs)

Receptor tryosine phosphatases are modular enzymes, containing a highly variable extracellular domain, a transmembrane domain and either one or two intracellular catalytic domains. When two catalytic domains are present, the C-terminal domain is usually catalytically inactive and possibly serves a regulatory function.

The extracellular domains are characterized by their variability between family members and the presence of domains which are often found in the extracellular domains of other transmembrane signalling proteins, such as the receptor tyrosine kinases and cell adhesion molecules. These domains include immunoglobulin-like domains and fibronectin type III repeats such as those found in cell adhesion molecules including NCAM and L1. These domains are suspected to mediate protein:protein interactions between RTP protein subunits, which interaction may modulate the signalling by the catalytic domains. RTPs can also contain extracellular domain motifs including MAM (meprin/A5/PTP mu) which is a conserved motif found in the metalloendoproteinase meprin A and B, RPTP mu and RPTP kappa, and the A5 neuronal cell surface glycoprotein, also known as neuropillin.

Examples of receptor tyrosine phosphatases include: receptor tyrosine phosphatase rho (GenBank ID No. AAF78091; mouse); protein tyrosine phosphatase receptor J (GenBank ID No. CAB94390; human); receptor-type tyrosine phosphatase D30 (GenBank ID No. NP_059032; rat); protein tyrosine phosphatase, receptor type, J (GenBank ID No. NP_058965; rat); protein tyrosine phosphatase, receptor type, C polypeptide associated protein (GenBank ID No. NP_058629; mouse); protein tyrosine phosphatase, receptor-type, T (dJ1121H13.2, GenBank ID No. CAB92429; human); receptor tyrosine phosphatase gamma (GenBank ID No. AAB25849; mouse); leukocyte-associated Ig-like receptor 1D isoform, LAIR-1D (GenBank ID No. AAF71275; human); and leukocyte-associated Ig-like receptor 1C isoform, LAIR-1C (GenBank ID No. AAF71274; human). Mutation sites for the modification of receptor tyrosine phosphatases may be predicted using the methods of the invention.

c. Serine/threonine Kinases

As the name implies, members of the receptor serine/threonine kinase family are characterized by a kinase domain that catalyzes the phosphorylation of serine and threonine residues. Members of the receptor serine/threonine kinase family are characterized by the presence of a ser/thr kinase domain, a transmembrane domain, and an extracellular domain. Receptor serine/threonine kinases include, for example, bone morphogenic protein receptor type IB (BMPR IB; GenBank ID No. NP_001194; human), bone morphogenic protein receptor type II (BMPR II; GenBank ID No. NP_001195; human), type I receptor Ser/Thr kinase (GenBank ID No 2211223A) and Myx (GenBank ID No. AAA91185; rat). Receptor serine/threonine kinases may be targeted for mutation using the methods of the invention.

3. Ligand-gated Ion Channels

Ligand-gated ion channels respond to the binding of agonist or ligand by opening or closing an ion channel. Such receptors are particularly prevalent in the mammalian nervous system, wherein the binding of neurotransmitters induces changes in membrane polarization. Other systems in which ligand-gated ion channels play important roles include the endocrine, immune, and cardiovascular systems, as well as in the regulation of cell growth. An online database of receptors in this superfamily is maintained at the Pasteur Institute (see Nicolas Le Novère and Jean-Pierre Changeux (1999). The Ligand-Gated Ion Channel database. *Nucleic Acid Research*, 27 : 340-342, incorporated herein by reference). There are three major superfamilies of extracellularly activated ligand-gated ion channel subunits, divided on the basis of the type of ligand that modulates receptor channel activity and by structural similarities.

The nicotinicoid receptors (nicotinic receptor, $GABA_A$ and $GABA_C$ receptors, glycine receptors, 5-HT_3 receptors and some glutamate activated anionic channels) are comprised of five homologous subunits.

The ATP-gated channels, also referred to as the P2X purinoceptors, form rapidly activated, non-selective cation channels and are activated by micromolar concentrations of extracellular ATP. At least seven P2X purinoceptor subtypes (P2X 1-7) have recently been identified and cloned (Buell *et al*, 1996, Eur. J. Neurosci. 8: 2221-2228), as have at least four human P2X receptor cDNAs, P2X1 (GenBank Accession No. X83688), P2X4 (GenBank Accession No. Y07684), P2X5a /P2X5b (GenBank Accession Nos. U49395 and U49396 respectively) and P2X7 (GenBank Accession No. Y09561).

The P2X purinoceptors comprise a family of glycosylated proteins, varying in size form 379 amino acids for P2X6 to 595 amino acids for P2X7 (Buell *et al*, 1996, supra). The DNAs encoding them exhibit high homology at the nucleotide level, and the amino acid sequence of these purinoceptors bears no significant homology with any other known receptor type. The P2X purinoceptors contain two predicted transmembrane domains (TM 1 and TM 2) with an intervening hydrophilic loop. The amino- and carboxy- termini are located intracellularly and a hydrophilic cysteine-rich loop resides extracellularly. The regularly spaced cysteines in this extracellular loop may form disulphide-bonded hairpin structures that stabilize the ligand binding pocket. There are three potential N-linked glycosylation sites also present on the hydrophillic

loop (Brake *et al*, 1994, Nature 371: 519-523). The P2X purinoceptors are comprised of three homologous subunits.

Finally, the glutamate activated cationic channels (NMDA receptors, AMPA receptors, Kainate receptors, etc.) are comprised of four homologous subunits.

4. Immune Recognition Receptors

Immune recognition receptors, such as the T cell receptors and surface IgM on immature B cells may also be targeted for mutation using the methods of the invention. Mutation data regarding specific sites that alter the signaling activity of a receptor that is closely related on a phylogenetic tree constructed as described herein may be applied in designing mutant immune recognition receptors.

B. Prediction of Receptor Sites to Mutate

The process involved for the prediction of sites to mutate according to the invention follows these general steps: 1) prepare a multiple sequence alignment of receptors in the same family as the receptor to be mutated; 2) perform phylogenetic profiling to place the receptor into a sub-family and thus determine the closest known relative(s) of the receptor; and 3) select sites to mutate based on mutant data available for relatives determined by phylogenetic profiling. Each of the necessary elements of this process is detailed below.

1. Multiple Sequence Alignment

Prior to the multiple sequence alignment, the receptor to be mutated must be placed into a family of related receptor proteins. This is performed by, for example, running a search using a sequence comparison program such as BLAST (Basic Local Alignment Search Tool; Altschul et al., 1990, J. Mol. Biol. 215: 403-410). The United States National Center for Biotechnology Information provides online access to the BLAST suite of programs at www.ncbi.nlm.nih.gov. The Blastp program, which searches protein databases with protein query sequences, is well suited to the task. Upon searching a database such as Swiss-Prot with the amino-acid sequence of the receptor to be mutated, the best scoring hits will be members of the same receptor family as the query sequence. Default parameters are acceptable for a search for this purpose. The "score" of the hits is judged by the "expect" or "E" value in the BLAST search. The lower the E value, the more similar the proteins are. The best scoring hits will have an E value between 0 and $1e^{-10}$.

For a rhodopsin class GPCR, all Blastp hits with an E value <0.1 will be other rhodopsin class family members.

Following the identification of the family to which the receptor of interest belongs, a high quality multiple amino acid sequence alignment is created. As used herein, the term "multiple sequence alignment" refers to an alignment of at least three receptor amino acid sequences such that amino acids in one sequence are aligned with similar amino acids in the other sequences. At least three amino acid sequences in a receptor family, preferably 4, 5, 7, 10, 15, 20, 30, 50, 100 or more sequences are used to generate the multiple sequence alignment. As used herein, the term "similar" when applied to amino acids means that one amino acid has similar size, charge, polar nature, lack or presence of aromatic groups, or functional qualities (e.g., the ability to serve as substrate for a kinase or the ability to participate in a particular secondary structure) as another amino acid. Similar amino acids include, for example, Met, Val, Leu and Ile; Ser and Thr; Asp, Glu and Asn; Gln, Lys and Arg; and Phe and Tyr. Greater numbers of sequences are preferred. For some receptors, such as those of the rhodopsin-like GPCR family, the level of overall sequence conservation, even within sub-families, is rather low, and reliable alignments can only be obtained for the transmembrane domains and the shorter of the loops between them. Manual, automated and semi-automated methods for multiple sequence alignment are well established. Manual methods are slow and labor intensive. Standard automated methods do not necessarily produce optimal alignments of, for example, GPCRs. A variety of such methods are available. The pileup program (GCG) and ClustalW (Thompson et. al. (1994) Nucleic Acids Res. 22, 4673-80) are the most widely used. Implementations of Hidden Markov Models (e.g. SAM (http://www.cse.ucsc.edu/research/compbio/sam.html) and diagonal (e.g Dialign (Morgenstern et. al. (1996) Bioinformatics 14, 290-294)) methods are also available. None of these methods successfully align conserved residues in diverse GPCRs at first pass. Future technical advances may make completely automated alignments that are sufficient for the method of the invention possible. Until then, however, a preferred embodiment of the invention uses a semi-automated method as described below.

To perform a semi-automated alignment, the first step is the prediction of transmembrane domains. This is done on the basis of hydropathy (using e.g. PEPTIDESTRUCTURE [GCG] , a program that implements the method of Kyte and Doolittle (1982, J. Mol. Biol. 157, 105-132). The prediction of transmembrane domains allows the next step, the manual identification and alignment of key conserved residues in the transmembrane regions and extracellular loops. In order to identify and align key conserved residues, one considers not only residues that are identical to those in similar positions of the transmembrane or loop domains, but also residues

that represent conservative substitutions. Conservative substitutions are those in which functionally or structurally similar amino acids occupy a given position in domains or sequences being compared.

As an example of the selection of key conserved residues, those used to anchor the alignment of GPCRs correspond to the following residues in the human alpha 1B adrenoceptor: Asn63 (an asparagine residue in TM1, typically preceded by Gly); Asp91 (an aspartate residue in TM2, typically preceded by Ala followed by a hydrophobic residue); C118 (a cysteine residue upstream of TM3 involved in a stabilizing disulfide bridge present in most rhodopsin-like GPCRs - where this is absent, the analogous residue is 25 residues N-terminal to the next residue described); Arg143 (in TM3; central to a highly conserved DRY motif); optionally Cys195 and Pro215 (residues between TM4 and 5 and in TM5 respectively, that cannot be identified unambiguously in all sequences); Pro309 (Proline residue in TM6, typically 2 residues after an aromatic amino-acid and followed by an aromatic amino-acid); Pro345 (proline residue in TM7, typically preceded by Asn). While these rules for identifying key residues are widely applicable, they are not universal. One of ordinary skill in the art is credited with the ability to assign appropriate residues by identification of conservative substitutions and interpretation of the context of residues.

Following the identification of conserved residues in the transmembrane regions that anchor the alignment, regions of sequence lying between the identified conserved residues may be aligned automatically using, for example, the *Insitu* option of PILEUP (GCG) (default gap opening and extension weights of 12 and 4 respectively give acceptable results). Figure 2 shows an example of a multiple sequence alignment of a small selection of GPCRs.

Phylogenetic Profiling

A phylogenetic profile is a statistical inference of the evolutionary relationship of one amino acid or nucleotide sequence to another or to various members of a related family. The aim of phylogenetic profiling is to allocate an orphan receptor to a sub-family of the larger super-family, on the basis of this evolutionary relationship. "Phylogenetic profiling" refers to the generation of a phylogenetic profile using the methods described herein or known in the art. A phylogenetic profile is most often represented by a "phylogenetic tree" which shows the evolutionary relationships of sequences to each other as the various branches on a tree.

To establish a phylogenetic profile for a group of receptors, a multiple sequence alignment of members of the family is first generated as detailed above. Positions in the multiple sequence alignment containing gaps in any of the sequences are excluded from the phylogenetic profile analysis as are regions where it is difficult to be confident in the reliability of the alignment, as is often the case in regions flanking gaps (see Beanland and Howe, 1992, Comp. Biochem. Physiol. 102B, 643-659. Phylogenetic trees are reconstructed using minimum evolutionary distance, maximum parsimony or other methods, as described below (reviewed by Nei, 1996, Ann. Rev. Genet. 30: 371-403; see also, Swofford et al., 1996, Phylogenetic inference, in *Molecular Systematics*, 2nd ed., ed. D. M. Hillis, C. Moritz, and B. K. Mable, Sinauer Associates, Sunderland, Massachusetts, pp. 407-514; Felsenstein, 1988, Ann. Rev. Genet. 22: 521-565; and Felsenstein, 1988, Ann. Rev. Ecol. Systemat. 19: 445-471, all of which are incorporated herein by reference). Figure 3a shows an example of the results of a phylogenetic analysis of GPCRs incorporating several orphan receptors. Figures 3b, c and d are enlarged views of different parts of the tree shown in Figure 3a.

1. Phylogenetic Profiling Methods Based on Evolutionary Distance

There are many methods for constructing phylogenetic trees based upon evolutionary distance, however the most commonly used distance methods are based on the principles of least squares and minimum evolution.

a. Least Squares-based Methods

In least squares methods, the minimum sum of the squared differences between observed pairwise distances and estimated pairwise distances for a given tree topology is computed. The tree topology with the smallest minimum sum of squared differences is chosen as the most likely to represent the actual tree. This method is described by Bulmer, 1991, Mol. Biol. Evol. 8: 868-883, which is incorporated herein by reference. Other methods based on least squares include those described by Cavalli-Sforza & Edwards, 1967, Am. J. Hum. Genet. 19: 122-257, and Fitch & Margoliash, 1967, Science 155: 279-284, both of which are incorporated herein by reference.

b. Minimum Evolution-based Methods

In minimum evolution methods, the branch lengths of a phylogenetic tree are estimated using an algorithm using pairwise distance data, and the total sum (S) of branch lengths is computed for each of the possible tree topologies. The topology that shows the smallest S value

is then selected as the most likely phylogenetic tree. This method is described in detail by Rzhetsky & Nei, 1992, Mol. Biol. Evol. 9: 945-967, and further defined by Rzhetsky & Nei, 1993, Mol. Biol. Evol. 10: 1073-1095, both of which are incorporated herein by reference. For The method of Kumar (1996, Mol. Biol. Evol. 13:584-593, incorporated herein by reference) is suggested for phylogenetic profiling involving more that 10 sequences; the method uses the neighbor joining tree method (see below) to extend a minimum evolution algorithm so that many sequences may be placed on a phylogenetic tree.

2. Phylogenetic Profiling Methods Based on Neighbor Joining

The neighbor joining method of phylogenetic profiling is a simplified version of the minimum evolution method of constructing a bifurcating phylogenetic tree (Saitou & Nei, 1987, Mol. Biol. Evol. 4: 406-425, incorporated herein by reference). In this approach, the S value is not calculated for all different possible topologies. Rather, the examination of different topologies is embedded in the profiling algorithm, such that only one tree is generated. The computation of S begins with a star phylogeny (all branches radiate from one point) in which all interior branches are assumed to have a length of zero, resulting in a value S_0. A new value, S_{ij}, is computed in which sequences *i* and *j* are paired and are separated from the remainder of the sequences that still form a star tree. If *i* and *j* are the neighbors joined by only one node, the S_{ij} is smaller than S_0. Therefore, computing S_{ij} for all pairs of sequences and choosing the smallest S_{ij}, one may identify a pair of neighbors. Once this one pair is identified, they are combined as a single unit and treated as a single sequence in the next step. This sequence is repeated until all multibranched nodes are resolved into bifurcating nodes, thereby constructing a phylogenetic tree.

3. Phylogenetic Profiling Methods Based on Likelihood

Adachi & Hasegawa (1995, *MOLPHY: Programs for Molecular Phylogenetics*, Tokyo Inst. Statist. Math., which is incorporated herein by reference) describe a maximum likelihood method of inferring a phylogenetic tree based on protein sequences. The method extends the empirical transition matrix for 20 different amino acids, established by Dayhoff et al. (1978, in *Atlas of Protein Sequence and Structure*, ed. M.O. Dayhoff, National Biomedical Research Foundation, Washington, DC, pp345-352, which is incorporated herein by reference), by using other transition matrices, including that of Jones et al. (1992, Comput. Appl. Biosci. 8: 275-282, incorporated herein by reference).

4. Phylogenetic Profiling Methods Based on Parsimony

In maximum parsimony methods, a given set of amino acid sequences is considered, and the amino acids of ancestral sequences for a hypothetical tree topology are inferred under the assumption that mutational changes occur in all directions among the 20 amino acids. The smallest number of amino acid substitutions that explain the entire evolutionary process for the given phylogenetic tree topology is then computed. This computation is then performed for all other topologies, and the tree topology that requires the smallest number of substitutions is chosen to be the best phylogenetic profile (Eck & Dayhoff, 1966, in *Atlas of Protein Sequences and Structure*, Natl. Biomed. Res. Found., Silver Springs, MD; Fitch, 1971, Sys. Zool. 20:406-416; Hartigan, 1973, Biometrics 29: 53-65, all of which are incorporated herein by reference).

5. Testing Inferred Phylogenetic Trees for Accuracy

There are a number of methods available for testing the accuracy of the trees generated using the various methods described above. For example, the interior branch test, described by Sitnikova et al. (1995, Mol. Biol. Evol. 12: 319-333, incorporated herein by reference), or Felsenstein's Bootstrap Test (Felsenstein, 1985, Evolution 39: 783-791, incorporated herein by reference) may be used to test the reliability of a tree generated according to the above-described methods. Felsenstein's Bootstrap Test uses a random resampling technique involving the construction of a new phylogenetic tree from a set of sequences obtained by random sampling with replacement. The random sampling with replacement and generation of a new tree is preformed many times, and the proportion of trees in which a given sequence cluster appears is computed. If the proportion is high (e.g., >0.95), the cluster is considered to be statistically significant. Felsenstein's Bootstrap Test works well when the sequences are more distantly related.

Selection of Residues for Mutagenesis

Known mutations leading to constitutive receptor activation, receptor inactivation and changes in agonist affinity are mapped onto a multiple sequence alignment. The phylogenetic profile of an orphan receptor is used to inform the predictive process. More specifically, the closest phylogenetic relative or relatives of a targeted receptor (that is, a receptor being targeted for mutation) are determined by the phylogenetic analysis, and mutation data available for those relatives are used to predict sites in the given receptor sequence to mutate. The "closest relative" of a given receptor protein is that which shares the most recent common ancestor on the

phylogenetic tree. It is the branch order on the tree, rather than the absolute evolutionary distance which determines relatedness in this instance.

Mutants leading to constitutive activation are predicted by reference to the sites mutated in the sub-family members for which data are available. Once a target receptor protein is allocated into a sub-family through the process of sequence alignment and phylogenetic profiling described herein, data regarding mutants of those sub-family members or close relatives may be readily retrieved from databases such as the MedLine database, or PubMed, either of which are available online. In addition, both U.S. and international patent databases may be readily searched to identify known mutants of a close relative that have a given phenotype. These databases are continually updated with publication information and may be searched by keywords to identify publications containing mutant data for a given receptor protein. In addition, there are databases maintained for mutation data from some specific types of receptors. For example, a database of GPCR mutation information is maintained online by the University of Texas Health Sciences Center at http://www.gcrdb.uthscsa.edu/GCR_Fam.html. Another example is the Ligand-gated ion channel receptor database maintained at the Pasteur Institute (Le Novère & Changeux, 1999, supra).

Once one or more sub-family members with mutant data available are identified, the mutant data are compared to identify corresponding sites to mutate in the target receptor. Mutation sites that do not generate a mutant with the desired activity (e.g., constitutive signaling activity) in all members of the set of relatives for which data are available are usually not considered in predicting sites to mutate. The term "usually" is used because it is possible, and even likely in some instances, that only a certain type of mutation will bring about the desired change in receptor activity. For example, the mutation of a single amino acid from an acidic to a neutral residue may have the desired effect, but mutation of the same site to a basic residue may not have the same effect in that receptor or one related to it. Therefore, it is important when comparing available mutation data for members of a given set of receptors, to note the type of amino acid change leading to the activity desired. For example, if the replacement of a basic amino acid with an acidic residue does not have the desired effect, but a mutation that alters the amino acid to a neutral residue does alter the activity as hoped, one has learned what type of change in a target receptor will be more likely to result in the desired change in activity of the target. The published information regarding mutants may thus be applied to predict the change to make in the target receptor.

In addition to the above-described methods of predicting sites for mutation in essentially any type of receptor, two mutation strategies, referred to elsewhere herein as Method 1 and Method 2, are applicable across a broad range of rhodopsin-family GPCR subtypes:

1. Mutation of the residue 16 residues N-terminal to the conserved proline, analogous to P309 in TM6 of the human alpha-1B adrenoceptor. This has lead to constitutive activity in all receptors tested. In a preferred embodiment of the invention, the residue is substituted with glutamate, however the substituent may be chosen from the set glutamate, glutamine, arginine, and lysine.

2. For chemokine, opioid/somatostatin, P2Y and related receptors, mutation of the conserved asparagine residue, analogous to N111 in TM3 of the type 1A angiotensin II receptor, to alanine leads to constitutive activity in three known examples.

The methods described herein may be readily applied to predicting any type of mutation for which data are available regarding other members of a given family, in both receptor and non-receptor proteins. Other receptor activities of interest include, for example, increases or decreases in agonist, inverse agonist or antagonist binding affinity, agonist or inverse agonist potency, cell surface expression, receptor stability, basal second messenger stimulation, and maximal agonist stimulated second messenger level. Non-receptor activities that can be targeted include, for example, protein:protein interaction, and enzyme activities such as protease activity, kinase activity, phosphatase activity, and lipase activity to name a very few.

Methods of Making Mutant Receptors Comprising Mutations Predicted According to the Invention.

A. Site-directed Mutagenesis

There are a number of site-directed mutagenesis methods known in the art which allow one to mutate a particular site or region in a straightforward manner. These methods are embodied in a number of kits available commercially for the performance of site-directed mutagenesis, including both conventional and PCR-based methods. Examples include the EXSITE™ PCR-based site-directed mutagenesis kit available from Stratagene (Catalog No. 200502; PCR based) and the QUIKCHANGE™ site-directed mutagenesis kit from Stratagene

(Catalog No. 200518; PCR based), and the CHAMELEON® double-stranded site-directed mutagenesis kit, also from Stratagene (Catalog No. 200509).

Older methods of site-directed mutagenesis known in the art relied upon sub-cloning of the sequence to be mutated into a vector, such as an M13 bacteriophage vector, that allows the isolation of single-stranded DNA template. In these methods one annealed a mutagenic primer (i.e., a primer capable of annealing to the site to be mutated but bearing one or more mismatched nucleotides at the site to be mutated) to the single-stranded template and then polymerized the complement of the template starting from the 3' end of the mutagenic primer. The resulting duplexes were then transformed into host bacteria and plaques were screened for the desired mutation.

More recently, site-directed mutagenesis has employed PCR methodologies, which have the advantage of not requiring a single-stranded template. In addition, methods have been developed that do not require sub-cloning. Several issues must be considered when PCR-based site-directed mutagenesis is performed. First, in these methods it is desirable to reduce the number of PCR cycles to prevent expansion of undesired mutations introduced by the polymerase. Second, a selection must be employed in order to reduce the number of non-mutated parental molecules persisting in the reaction. Third, an extended-length PCR method is preferred in order to allow the use of a single PCR primer set. And fourth, because of the non-template-dependent terminal extension activity of some thermostable polymerases it is often necessary to incorporate an end-polishing step into the procedure prior to blunt-end ligation of the PCR-generated mutant product.

The protocol described below accommodates these considerations through the following steps. First, the template concentration used is approximately 1000-fold higher than that used in conventional PCR reactions, allowing a reduction in the number of cycles from 25-30 down to 5-10 without dramatically reducing product yield. Second, the restriction endonuclease DpnI (recognition target sequence: 5-Gm6ATC-3, where the A residue is methylated) is used to select against parental DNA, since most common strains of *E. coli* Dam methylate their DNA at the sequence 5'-GATC-3'. Third, Taq Extender is used in the PCR mix in order to increase the proportion of long (i.e., full plasmid length) PCR products. Finally, Pfu DNA polymerase is used to polish the ends of the PCR product prior to intramolecular ligation using T4 DNA ligase. The method is described in detail as follows:

PCR-based Site Directed Mutagenesis

Plasmid template DNA (approximately 0.5 pmole) is added to a PCR cocktail containing: 1x mutagenesis buffer (20 mM Tris HCl, pH 7.5; 8 mM $MgCl_2$; 40 ug/ml BSA); 12-20 pmole of each primer (one of skill in the art may design a mutagenic primer as necessary, giving consideration to those factors such as base composition, primer length and intended buffer salt concentrations that affect the annealing characteristics of oligonucleotide primers; one primer must contain the desired mutation, and one (the same or the other) must contain a 5' phosphate to facilitate later ligation), 250 uM each dNTP, 2.5 U Taq DNA polymerase, and 2.5 U of Taq Extender (Available from Stratagene; See Nielson et al. (1994) Strategies 7: 27, and U.S. Patent No. 5,556,772). The PCR cycling is performed as follows: 1 cycle of 4 min at 94°C, 2 min at 50°C and 2 min at 72°C; followed by 5-10 cycles of 1 min at 94°C, 2 min at 54°C and 1 min at 72°C. The parental template DNA and the linear, PCR-generated DNA incorporating the mutagenic primer are treated with DpnI (10 U) and Pfu DNA polymerase (2.5U). This results in the DpnI digestion of the in vivo methylated parental template and hybrid DNA and the removal, by Pfu DNA polymerase, of the non-template-directed Taq DNA polymerase-extended base(s) on the linear PCR product. The reaction is incubated at 37°C for 30 min and then transferred to 72°C for an additional 30 min. Mutagenesis buffer (115 ul of 1x) containing 0.5 mM ATP is added to the DpnI-digested, Pfu DNA polymerase-polished PCR products. The solution is mixed and 10 ul are removed to a new microfuge tube and T4 DNA ligase (2-4 U) is added. The ligation is incubated for greater than 60 min at 37°C. Finally, the treated solution is transformed into competent *E. coli* according to standard methods.

B. Expression of Mutant Receptors

Mutant receptors are expressed in cells and evaluated for desired alterations in activity following introduction of a nucleic acid encoding the mutant receptor to the cells. Nucleic acids may be introduced to cells in any of a number of ways that are well established in the art. These include, for example, lipofection, electroporation, calcium phosphate precipitation, and viral transduction. These methods are detailed, for example, in Current Protocols in Molecular Biology (Ausubel et al., 1987, John Wiley & Sons).

i. Vectors:

Vectors useful for the introduction, maintenance and expression of foreign gene sequences in eukaryotic cells are well known in the art. Plasmids, cosmids, episomal vectors,

and viral vectors, such as retrovirus-based and adenovirus-based vectors are widely available and a vector suitable for a given cell type or application may be readily chosen by one of skill in the art. The attributes required of a vector useful in the practice of the invention include a promoter or promoter/enhancer combination that is active in the intended cell type. The promoter may be constitutively active, or it may be regulated by the presence or absence of an inducing agent, such as temperature change or tetracycline. If the vector is to be expanded in bacteria, it should have a bacterial or bacteriophage origin of replication. It is advantageous if eukaryotic expression vectors encode at least one intron, and eukaryotic expression vectors also generally contain a polyadenylation signal sequence 3' of the coding sequence. Vectors useful in the practice of the invention will usually contain sequences encoding a selectable marker, although co-transfection with a separate selectable marker plasmid may also be practiced.

ii. Cells:

Cells useful for the expression and evaluation of mutant receptor activities are most often, but not necessarily, eukaryotic, and may be derived from, for example, yeasts such as S. cerevisiae or other fungal cells, insect cells, or mammalian cells, including human cells. Any cell type that can be transfected, including, for example, NIH3T3, CHO, COS-7, HeLa, is suitable for evaluation of receptor activity. Cells may naturally express the wild-type receptor. However, it is preferred that cells used to evaluate the effects of a mutation predicted using the methods of the invention do not express the wild-type form of the receptor being mutated. Comparisons of receptor activities may then be made by transfecting cells separately with the wild-type and mutant receptors and monitoring downstream signalling.

iii. Lipofection:

Lipofection reagents and methods suitable for transient transfection of a variety of transformed and non-transformed or primary cells are widely available, making lipofection an attractive method of introducing constructs to eukaryotic, and particularly mammalian cells in culture. For example, LipofectAMINE™ (Life Technologies) or LipoTaxi™(Stratagene) transfection kits are commercially available. Other companies offering reagents and methods for lipofection include Bio-Rad Laboratories, CLONTECH, Glen Research, InVitrogen, JBL Scientific, MBI Fermentas, PanVera, Promega, Quantum Biotechnologies, Sigma-Aldrich, and Wako Chemicals USA.

Lipofection is performed essentially as described in the early publication by Felgner et al., 1987, supra. Numerous variations on the method and on the cationic lipid materials themselves are known and have been described. Generally, nucleic acids are mixed with

suspensions of cationic lipids to form complexes of nucleic acid and lipid that may fuse with and penetrate the plasma membrane of a cell. The complexes are generally formed and applied to cells in either the absence or a reduced amount of serum (e.g., about 2 to 0.5% or less). Complexes are generally contacted with the cells for times that vary from 30 minutes to overnight, followed by rinsing and replacement of serum-containing medium, if desired. The time of exposure of cells to lipid/nucleic acid complexes and the relative and absolute amounts of lipid and nucleic acid to use are empirically determined through routine experimentation. Generally, a series of trial transfections is performed, each with varying amounts of the transfection reagents and varying amounts of cell exposure time. Lipofection tends to kill a proportion of the cells, however, the dose of lipid and DNA and the time of exposure with the lowest mortality will not necessarily yield the highest transfection efficiency (i.e., the highest number of cells taking up and expressing a marker). One of skill in the art may readily determine the lipofection conditions that yield the highest efficiency transfection. Selection for transfectants that stably maintain the exogenous construct may be performed beginning approximately 12 to 48 hours after lipofection.

iv. Electroporation:

Electroporation is also well known in the art as a means of introducing nucleic acid constructs to cells. Exemplary electroporation conditions are presented below. For this method, exponentially growing or early stationary phase cells are trypsinized and prepared as a single cell suspension. An aliquot of the cell suspension is removed for counting, and the remaining cells are subjected to centrifugation. The supernatant is aspirated and the pellet is resuspended in 5 ml of electroporation buffer (20 mM HEPES pH 7.3, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose). The cells are recentrifuged, the supernatant aspirated, and the cells resuspended in electroporation buffer containing 1 mg/ml acetylated bovine serum albumin. The final cell suspension contains approximately 3 x 10^6 cells/ml. Electroporation should be performed immediately following this resuspension.

Supercoiled plasmid DNA is added to a sterile cuvette with a 0.4 cm electrode gap (Bio-Rad). The final DNA concentration is generally at least 120 μg/ml. 0.5 ml of the cell suspension (containing approximately 1.5 x 10^6 cells) is then added to the cuvette, and the cell suspension and DNA solutions are gently mixed. Electroporation is performed with a Gene-Pulser apparatus (Bio-Rad). Capacitance and voltage are set at 960 μF and 250-300 V, respectively. As voltage increases, cell survival decreases, but the percentage of surviving cells that stably incorporate the introduced DNA into their genome increases dramatically. Given these parameters, a pulse time of approximately 14-20 msec should be observed.

Electroporated cells are maintained at room temperature for approximately 5 min, and the contents of the cuvette are then gently removed with a sterile transfer pipette. The cells are added directly to prewarmed growth medium and incubated as appropriate for that cell type. The following day, the medium is aspirated and replaced with fresh medium and cells are incubated for a further 16-48 hrs before treatment with an appropriate selective agent, such as G418, in order to select cells that stably maintain the mutant construct.

v. Calcium Phosphate-mediated Transfection:

The method of calcium phosphate precipitation is described by, for example, Chen and Okayama, 1988, BioTechniques 6: 632-638, which is incorporated herein by reference. The approach is one of the oldest methods of eukaryotic gene transfer, and has a low efficiency compared, for example, to lipofection or electroporation, yet the method is very well characterized and reliable.

vi. Viral-mediated Gene Transfer:

Viral vectors used for gene transfer are frequently derived from retroviruses, but may be derived from other classes of viruses, for example, adenoviruses, adeno-associated virus, Epstein-Barr virus, papilloma viruses or paramyxoviruses. Viral vectors useful in the methods of the invention are preferably, but not necessarily, replication defective.

Retroviral vectors in general are described by Miller et al., 1989, BioTechniques 7: 980-990, and cell lines capable of supplying the viral proteins necessary for packaging of replication-defective retroviral vectors in *trans* are described by Miller, 1990, Human Gene Ther. 1: 5-14.

vii. Selectable markers:

Due to the relative infrequency of genomic integration events that generate functional transgenes, the isolation of cells that have stably integrated a given mutant receptor construct is generally not favored unless a selectable marker is also expressed by the construct (selectable markers can select for cells with episomally-maintained transgene constructs as well). Selectable markers can include, for example, positive selection markers, which permit a cell carrying and expressing the marker to survive treatment with a selective agent, or genes encoding fluorescent marker polypeptides (e.g., Green Fluorescent Protein, GFP). Among the positive selectable markers are the neomycin phosphotransferase, xanthine-guanine phosphoribosyl transferase, dihydrofolate reductase, adenosine deaminase, CAD (carbamyl phosphate synthase, aspartate transcarbamylase, dihydro-orotase), mdr1 (multidrug resistance 1) and glutamine synthase genes, as well as the genes for hygromycin, puromycin, and Histidine D resistance. The dose of selective agent differs with the choice of agents, but generally one aims for a dose that kills non-expressing cells in about 5 to 7 days and allows the emergence of colonies of resistant cells over

about 7 to 14 days. When a fluorescent polypeptide is used as a selectable marker, fluorescence-activated cell sorting (FACS) is performed to identify and isolate cells that stably express the marker. FACS methods are well known in the art.

Methods of Evaluating the Effect of Mutations Predicted According to the Invention

The functional effect of a receptor mutation predicted using a method of the invention may be determined in any of a number of ways, depending upon the exact receptor mutated and the normal signaling pathways used by that receptor. Assays for alterations in ligand binding may be used, for example, as may assays to detect changes in downstream signalling pathway activity. The assay selected will also depend upon the receptor function one wishes to alter (e.g., constitutive activation of downstream signaling, enhanced agonist binding affinity, enhanced agonist potency, etc.). Assays include, for example, ligand binding assays, GTPase activation/GTP binding assays, adenylate cyclase assays, cAMP assays, assays for PI breakdown and DAG production, assays for IP3 levels, assays for PLC-γ-activity, PKC activation assays, membrane polarization assays (Patch clamp), tyrosine kinase assays, MAP kinase assays, and reporter gene-based assays. For each assay, a comparison of activity of wild-type versus mutant receptors allows a determination of the actual effect of the mutation predicted according to the methods of the invention.

A. Ligand Binding Assays:

When a ligand is known, one may use receptors expressed on a cell or use isolated membranes to evaluate the binding characteristics of labeled (e.g., radiolabeled or fluorescently labeled) ligand by a mutant receptor relative to binding of ligand by cells or membranes containing wild-type receptor. The relative kinetics of binding may be evaluated, for example, by monitoring the displacement of labeled ligand by known quantities of unlabeled ligand on isolated membranes. For constitutively active GPCRs, an increase in agonist binding affinity compared with the wild type receptor can be used as a marker of constitutive activation.

B. GTPase/GTP Binding Assays:

For GPCRs, a measure of receptor activity is the binding of GTP by cell membranes containing receptors. In the method described by Traynor and Nahorski, 1995, Mol. Pharmacol. 47: 848-854, incorporated herein by reference, one is essentially measuring G-protein coupling to membranes. Membranes isolated (using methods known in the art) from cells which express the receptor are incubated in a buffer containing 20 mM HEPES, pH 7.4, 100 mM NaCl, and 10

mM MgCl2, 80 pM ^{35}S-GTPγS and 3 μM GDP. The assay mixture is incubated for 60 minutes at 30°C, after which unbound labeled GTP is removed by filtration onto GF/B filters. Bound labeled GTP is measured by liquid scintillation counting.

GTPase activity is measured by incubating the membranes containing the mutant receptor with γ^{32}P-GTP. Active GTPase will release the label as inorganic phosohate, which is detected in the supernatant by scintillation counting.

C. Downstream Pathway Activation Assays:

1. Adenylate Cyclase Assay:

Assays for adenylate cyclase activity are described by Maenhault et al., 1990, Biochem. Biophys. Res. Comm. 173: 1169-1178. Briefly, membranes containing approximately 50 to 75 μg of protein are incubated with a reaction medium containing 65 mM sucrose, 5 mM phosphocreatine, 10 U/ml creatine kinase, 0.04% BSA, 50 mM Tris, pH 7.4, 5 mM MgCl2, 0.25 mM EDTA, 0.12 mM RO20-1724 (phosphodiesterase inhibitor), 0.1 mM ATP, 0.1 mM GTP, and between 1.5 to 2.5 μCi/sample of α^{32}P-ATP. Following the addition of the membrane mixture, the assay is incubated at 31°C for 1 hour.

2. cAMP Assay:

Intracellular or extracellular cAMP is measured using a cAMP radioimmunoassay (RIA) or cAMP binding protein according to methods widely known in the art. For example, Horton & Baxendale, 1995, Methods Mol. Biol. 41: 91-105, which is incorporated herein by reference, describes an RIA for cAMP.

A number of kits for the measurement of cAMP are commercially available, such as the High Efficiency Fluorescence Polarization-based homogeneous assay marketed by LJL Biosystems and NEN Life Science Products.

3. Phospholipid breakdown, DAG production and Inositol Triphosphate levels:

Receptors that activate the breakdown of phospholipids may be monitored for changes due to mutations predicted using the methods of the invention by monitoring either phospholipid breakdown, DAG production or Inositol triphosphate (IP_3) levels. Methods of measuring each

of these are described in **Phospholipid Signaling Protocols**, edited by Ian M. Bird. Totowa, NJ, Humana Press, 1998, which is incorporated herein by reference.

An assay suitable to be adapted for monitoring mutant-receptor activated PI hydrolysis is also described by Sevva et al., 1986, Biochem. Biophys. Res. Comm. 140:160-166 and Peralta et al.,1988, Nature 334:434-437. Briefly, the functional assay involves labeling of cells with ^{3}H-myoinositol for at least 48 hours. Following incubation with labeled myoinositol, the cells are lysed and the suspension is extracted with 3 ml of $CHCl_3$/ MeOH (1:1). After centrifugation (3200 rpm for 5 min), the upper aqueous phase is removed and diluted with 2 ml H_2O and centrifuged again. The supernatants are loaded on columns containing 1 ml Dowex 1x8 AG resin previously equilibrated with 5 mM myoinositol and washed with 9 ml of 5 mM myoinositol followed by 8 ml of 60 mM sodium formate, 5 mM sodium borate. All of the inositol phosphates (IP1, IP2, IP3) are eluted together with 6 ml of 0.1 M formic acid, 1M ammonium formate. 3 ml of the eluates are removed and counted with 20 ml scintillation fluid for analysis.

4. PKC activation assays:

Growth factor receptor tyrosine kinases tend to signal via a pathway involving activation of Protein Kinase C (PKC), which is a family of phospholipid- and calcium-activated protein kinases. The PKC activation ultimately results in the transcription of an array of proto-oncogene transcription factor-encoding genes, including c-fos, c-myc and c-jun, proteases, protease inhibitors, including collagenase type I and plasminogen activator inhibitor, and adhesion molecules, including intracellular adhesion molecule I (ICAM I). Assays designed to detect increases in gene products induced by PKC may be used to monitor PKC activation and thereby receptor activity. In addition, the activity of receptors that signal via PKC may be monitored through the use of reporter gene constructs driven by the control sequences of genes activated by PKC activation. This type of reporter gene-based assay is discussed in more detail below.

5. Membrane polarization assays for measurement of receptor activity:

Electrophysiological measurements of receptor activity may be performed using standard patch-clamp techniques, as described by Hoo et al. (1994, Receptors and Channels 2: 327), and summarized as follows.

Electrophysiological recordings are performed in a standard extracellular solution composed of 140 mM NaCl, 5.4 mM KCl, 1.0 mM $MgCl_2$, 1.3 mM $CaCl_2$, 5.0 mM HEPES and glucose to an osmolarity of 300 mOsm and pH adjusted to 7.2 with 1 mM NaOH.

For ion permeability studies, two other recording solutions are used, including a low calcium solution (140 mM NaCl, 1.0 mM MgCl2, 5.0 mM HEPES (pH 7.2 with NaOH)), and a low sodium solution (110 mM CaCl2, 1.0 mM MgCl2, 5.0 mM HEPES (pH to 7.2 with $Ca(OH)_2$).

Electrodes are constructed from thin-walled borosilicate glass (WPI Instruments), pulled to a fine point (1-2 μm in width) and filled with an intracellular solution composed of 140 mM CsCl, 1.0 mM MgCl2, 10 mM EGTA, 10 mM HEPES with pH adjusted to 7.2 with Cs(OH)2 and an osmolarity of 300 mOsm.

Whole cell voltage clamp recordings are carried out using an Axopatch 1B amplifier (Axon Instruments) or its equivalent. Agonists and antagonists are rapidly perfused over the cells through a multibarrel array of square glass tubes, the position of which is adjusted using a piezomotor under computer control. With this system it is possible to rapidly exchange solutions flowing over the cell and thus carry out extensive studies of receptor pharmacology.

6. Kinase assays:

Assays for the activity of other signal transduction pathways regulated by a given receptor protein are known in the art. For example, direct assays for tyrosine kinase activity using known synthetic or natural tyrosine kinase substrates and labeled phosphate are well known, as are similar assays for other types of kinases (e.g., Ser/Thr kinases).

7. Transcriptional reporters for downstream pathway activation:

The intracellular signal that is transduced is generally initiated by the specific interaction of an extracellular signal, e.g., a ligand or agonist, with a receptor or ion channel present on the cell surface. This interaction sets in motion a cascade of intracellular events, the ultimate consequence of which is a rapid and detectable change in the transcription or translation of a gene. A mutation predicted using the methods of the invention will preferably have the same effect on downstream signalling as binding of agonist by the wild-type receptor. As used herein "promoter" refers to the transcriptional control elements necessary for receptor-mediated regulation of gene expression, including not only the promoter, but also any enhancers or transcription-factor binding sites necessary for receptor-regulated expression. By selecting promoters that are responsive to the intracellular signals resulting from agonist binding or an activating mutation, and operatively linking the selected promoters to reporter genes whose

transcription, translation or ultimate activity is readily detectable and measurable, the transcription based reporter assay provides a rapid indication of whether a specific receptor or ion channel is activated.

Reporter genes such as luciferase, CAT or β-galactosidase are well known in the art, as are assays for detection of their products.

Transcription-based reporter assays can be used to test functional ligand-receptor or ligand-ion channel interactions for categories of cell surface-localized receptors including, but not limited to ligand-gated ion channels and voltage-gated ion channels, G protein-coupled receptors and growth factor receptors. Examples of each group include:

a) ligand-gated ion channels: nicotinic acetylcholine receptors, GABA (gamma-aminobutyric acid) receptors, excitatory receptors (e.g., glutamate and aspartate), and the like;

b) voltage-gated ion channels: calcium channels, potassium channels, sodium channels, NMDA receptor (actually a ligand-gated, voltage-dependent ion channel) and the like;

c) G protein-coupled receptors: adrenergic receptors, muscarinic receptors and the like.

d) Growth factor receptors (Both RTKs and non-RTKs): Nerve growth factor NGF, heparin binding growth factors and other growth factors.

Transcriptional control elements include, but are not limited to, promoters, enhancers, and repressor and activator binding sites. Suitable transcriptional regulatory elements may be derived from the transcriptional regulatory regions of genes whose expression is rapidly induced, generally within minutes, of contact between the cell surface protein and the effector protein that modulates the activity of the cell surface protein. Examples of such genes include, but are not limited to, the immediate early genes (see, Sheng et al. (1990) Neuron 4: 477-485), such as c-fos. Immediate early genes are genes that are rapidly induced upon binding of a ligand to a cell surface protein. The induction of immediate early gene transcription does not require the synthesis of new regulatory proteins. The transcriptional control elements that are preferred for use in the gene constructs include transcriptional control elements from immediate early genes, elements derived from other genes that exhibit some or all of the characteristics of the immediate early genes, or synthetic elements that are constructed such that genes in operative linkage therewith exhibit such characteristics. The characteristics of preferred genes from which the transcriptional control elements are derived include, but are not limited to, low or undetectable expression in quiescent cells, rapid induction at the transcriptional level within minutes of extracellular simulation, induction that is transient and independent of new protein synthesis, subsequent shut-off of transcription requires new protein synthesis, and mRNAs transcribed from these genes have a short half-life. It is not necessary for all of these properties to be present.

One gene that is responsive to a number of different stimuli is the c-fos proto-oncogene. The c-fos gene is activated in a protein-synthesis-independent manner by growth factors, hormones, differentiation-specific agents, stress, and other known inducers of cell surface proteins. The induction of c-fos expression is extremely rapid, often occurring within minutes of receptor stimulation. This characteristic makes the c-fos regulatory regions particularly attractive for use as a reporter of receptor activation.

It is known in the art which receptors activate c-fos expression. One may determine whether an orphan receptor or an uncharacterized receptor activates the c-fos or other regulatory sequences by co-transfecting the wild-type receptor, under control of a strong promoter such as the CMV promoter/enhancer, with a c-fos (or other) reporter construct and measuring expression of the reporter in comparison to cells transfected with reporter alone (or with reporter and an expression vector minus receptor sequences). Generally, the overexpression of a receptor, even in the absence of ligand, will result in some signal transduction by the receptor.

The c-fos regulatory elements include (see, Verma et al., 1987, Cell 51: 513-514): a TATA box that is required for transcription initiation; two upstream elements for basal transcription, and an enhancer, which includes an element with dyad symmetry and which is required for induction by TPA, serum, EGF, and PMA.

The 20 bp transcriptional enhancer element located between -317 and -298 bp upstream from the c-fos mRNA cap site, is essential for serum induction in serum starved NIH 3T3 cells. One of the two upstream elements is located at –63 to -57 and it resembles the consensus sequence for cAMP regulation.

The transcription factor CREB (cyclic AMP responsive element binding protein) is, as the name implies, responsive to levels of intracellular cAMP. Therefore, the activation of a receptor that signals via modulation of cAMP levels may be monitored by measuring either the binding of the transcription factor, or the expression of a reporter gene linked to a CREB-binding element (termed the CRE, or cAMP response element). The DNA sequence of the CRE is TGACGTCA. Reporter constructs responsive to CREB binding activity are described in U.S. Patent No. 5,919,649.

A CREB-responsive reporter construct is transfected into cells with wild-type or mutant receptor expression, and the relative level of receptor activity is determined by the reporter activities. Alternatively, the binding of CREB to DNA may be monitored using, for example, the well known electrophoretic mobility shift assay.

Other promoters and transcriptional control elements, in addition to the c-fos elements and CREB-responsive constructs, include the vasoactive intestinal peptide (VIP) gene promoter (cAMP responsive; Fink et al., 1988, Proc. Natl. Acad. Sci. 85:6662-6666); the somatostatin gene promoter (cAMP responsive; Montminy et al., 1986, Proc. Natl. Acad. Sci. 8.3:6682-6686); the proenkephalin promoter (responsive to cAMP, nicotinic agonists, and phorbol esters; Comb et al., 1986, Nature 323:353-356); the phosphoenolpyruvate carboxy-kinase (PEPCK) gene promoter (cAMP responsive; Short et al., 1986, J. Biol. Chem. 261:9721-9726); the NGFI-A gene promoter (responsive to NGF, cAMP, and serum; Changelian et al., 1989, Proc. Natl. Acad. Sci. 86:377-381); and others that may be known to or prepared by those of skill in the art.

Screening Assays Using Proteins Comprising Mutations Predicted According to the Invention

Mutant proteins (e.g., receptor mutants) made according to the prediction methods described herein are useful as the basis of screening assays to identify modulators of the wild-type protein. Compounds found to modulate the activity of a receptor may be, for example, agonists, inverse agonists, partial agonists, or may have other modulatory function. Methods of screening for modulators of protein activity (e.g., receptor activity) using mutant proteins comprise the steps of a) contacting a candidate compound with the mutant protein, and b) determining the activity of the protein in the presence of the candidate compound relative to its activity in the absence of the candidate compound.

Methods of producing a mutant receptor or other protein comprising a mutation predicted according to the methods of the invention are described herein above and are also well known to those skilled in the art. Methods of determining the activity of a receptor are described in the preceding section "Methods of Evaluating the Effect of Mutations Predicted According to the Invention". For non-receptor proteins, activities (e.g., enzyme activities) may be measured according to appropriate methods as known in the art. As used herein, the terms "candidate compound" and "candidate modulator" refer to a composition being evaluated for the ability to modulate the activity of a receptor or other protein. The term "modulate the activity" means that a compound causes a change in the activity of a given receptor or other protein, wherein the activity may be signaling activity in vivo, enzyme activity, or any of the activities described in the preceding section of this specification. For example, a change may occur in any of ligand binding, GTPase/GTP binding, or downstream pathway activity(ies). It is preferred that a modulator found to induce a change in a receptor activity in vitro has an effect on the signaling activity of the receptor (mutant or wild-type) in vivo. The change induced by a modulator may

be an increase or a decrease in an activity. It is preferred, although not required, that a modulator be an inverse agonist, an agonist or a partial agonist.

The candidate modulator may be a synthetic compound, a mixture of compounds, or may be a natural product (*e.g.* a plant extract or culture supernatant).

Candidate modulators can be screened from large libraries of synthetic or natural compounds. Numerous means are currently used for random and directed synthesis of saccharide, peptide, and nucleic acid based compounds. Synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH), and Microsource (New Milford, CT). A rare chemical library is available from Aldrich (Milwaukee, WI). Combinatorial libraries are available and can be prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available from e.g., Pan Laboratories (Bothell, WA) or MycoSearch (NC), or are readily produceable by methods well known in the art. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

Useful compounds may be found within numerous chemical classes. Useful compounds may be organic compounds, or small organic compounds. Small organic compounds have a molecular weight of more than 50 yet less than about 2,500 Daltons, preferably less than about 750, more preferably less than about 350 daltons. Exemplary classes include heterocycles, peptides, saccharides, steroids, and the like. The compounds may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways to enhance their stability, such as using an unnatural amino acid, such as a D-amino acid, particularly D-alanine, by functionalizing the amino or carboxylic terminus, e.g. for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like.

Candidate modulators will be effective at varying concentrations, depending upon the nature of the compound and upon the nature of its interaction with the receptor or other protein. Therefore, candidate modulators should be screened at varying concentrations. A concentration range of about 10 mM to about 1 fM should be used for screening, preferably about 1 mM to 1 pM, more preferably about 1 μM to about 1 pM. The association constants of agents that bind and/or inhibit protein activities will generally in the range of about 1 mM to about 1 fM, and optimally in the range of about 1 μM to about 1 pM or less.

The contacting step in the screening methods according to the invention will vary depending upon the exact assay format being used. For example, when the mutant protein is produced in cultured cells, the contacting may comprise adding the candidate modulator to the culture medium. Alternatively, in in vitro assays, for example, those using purified mutant protein, the contacting step will comprise adding or applying a solution containing the candidate modulator to the assay reaction. Depending upon the selected assay, the mutant protein may be exposed to the candidate modulator before the mutant protein is contacted with other assay reagents, or the candidate modulator may be added after any or all other components of the assay.

EXAMPLES

Example 1: Prediction of mutations which constitutively activate GPR8

The method of the invention was used to predict sites for mutagenesis in the orphan G-protein coupled receptor GPR8. The key conserved residues used to assist in construction of the multiple sequence alignment were: N63, D91, C117, R142, C197, P275 and P313. A phylogenetic tree was constructed showing to establish the position of this receptor within the Rhodopsin-type GPCR family (Figure 3). This phylogenetic profile analysis showed GPR8 to be a member of the opioid/somatostatin group of receptors. Mutants of the delta-opioid receptor that show constitutive activity have been constructed (Befort et al. (1999) J. Biol. Chem. 274, 18574-18581). The mutants D128A, D128N, Y129F, Y129A and Y308F of this receptor all showed some degree of enhanced constitutive activity, with D128A having the highest activity. A second multiple sequence alignment was constructed incorporating GPR8 and other members of the opioid/somatostatin group of receptors (Figure 4). From this alignment we deduced that D124 of GPR8 corresponds to D128 of the delta-opioid receptor. We would therefore mutate D124 to alanine in GPR8. In addition the residues described in Method 1 & Method 2 herein above can be identified in GPR8. N127 is equivalent to N111 of the type II angiotensin receptor (Figure 5) and we would therefore mutate N127 to alanine in GPR8. T259 is equivalent to A293 of the alpha-1B-adrenoceptor (figure 5) and we would therefore mutate T259 to glutamate in GPR8. A similar strategy is applicable to GPR7, with the mutations D116A, N119A and T250E being indicated.

Example 2: Prediction of mutations which constitutively activate GPR10

Phylogenetic profiling showed that GPR10 is closely related to the neuropeptide Y and neurokinin receptors (Figure 3). A mutation Y216E in TM5 of the neurokinin I receptor leads to increased agonist affinity. An analogous residue (Y244) is identifiable in GPR10 (Figure 6). V247 in GPR10 is analogous to the residue in 1) above. The method of the invention predicts that one should make separate mutations of Y244 to glutamate and V247 to glutamate in order to constitutively activate the receptor.

Example 3: Prediction of mutations which constitutively activate GPR1

Phylogenetic profiling showed that GPR1 is related to the fMET peptide receptors and the C5A anaphylotoxin receptor (Figure 3). A mutation R175A in extracytoplasmic loop 2 of the last leads to increased agonist affinity. An analogous residue (R176) is identifiable in GPR1 (Figure 7). F245 in GPR1 is analogous to the residue identified in Method 1 and N120 is analogous to the residue identified in Method 2. These methods predict that one should separately mutate R176 to alanine, F245 to glutamate and N120 to alanine in order to constitutively activate the receptor.

Example 4: Prediction of mutations which constitutively activate GPR17

Phylogenetic profiling showed that GPR17 is a member of the P2Y-like receptor sub-family (Figure 3). The residues described in Method 1 & Method 2 above are identifiable in this receptor (Figure 8). The method of the invention predicts that one should separately mutate N114 to alanine and V234 to glutamate. A similar strategy is applicable to several other orphans. For example, the method predicts that one should mutate as follows: GPR7 (N119A, T250E), GPR4 (N100A, K223E), GPR15 (N116A, I240E), GPR20 (N133A, M240E), HM74 (N110A, I230E), OGR1 (N104A, Q227E), EBI2 (N114A, L243E), BONZO (N112A, L230E), RDC1 (N127A, R251E), O15218 (N136A, C257E), H963 (N97A, L222E), GPR30 (N140A, L258E).

Example 5: Prediction of mutations which constitutively activate GPR21

Phylogenetic profiling showed that GPR21 is not specifically related to any major receptor subtype (Figure 3). Only the residue described in Method 1 above can be identified (Figure 9). This Method indicates that one should mutate A251 to glutamate. This strategy is applicable to a number of other orphan receptors, e.g. GPR2 (L238E), GPR5 (V224E), GPR13 (I230E), GPR18 (I231E), GPR22 (F312E), GPR25 (L242E), GPR31 (Q221E), GPR38 (V297E),

GPR39 (I282E), GPR40 (A223E), GPR41 (A224E), GPR42 (A224E), GPR43 (V221E), MGR (Y263E).

Table 1

GPCR	Abbreviation	Swiss-Prot I.D.	Accession number
Alpha-IB adrenoceptor	AlAB	AIAB_HUMAN	P35368
Angiotensin 2 receptor I	AG2R	AG2R_HUMAN	P30556
APJ	-	APJ_HUMAN	P35414
BONZO	BONZ	BONZ_HUMAN	O00574
C5-Anaphylotoxin receptor	C5AR	C5AR_HUMAN	P21730
EBV-induced receptor 2	EB12	EB12_HUMAN	P32249
fMet peptide receptor 1	FML1	FML1_HUMAN	P25090
Gastrin receptor	GASR	GASR_HUMAN	P32239
GPR-9-6	GC96	GC96_HUMAN	P51686
GPRI	GPRI	GPRI_HUMAN	P46091
GPR2	GPR2	GPR2_HUMAN	P46092
GPR4	GPR4	GPR4_HUMAN	P46093
GPR5	GPR5	GPR5_HUMAN	P46094
GPR6	GPR6	GPR6_HUMAN	P46095
GPR7	GPR7	GPR7_HUMAN	P48145
GPR8	GPR8	GPR8_HUMAN	P48146

GPR10	GPRA	GPRA_HUMAN	P49683
GPR12	GPRC	GPRC_HUMAN	P47775
GPR13	GPRD	GPRD_HUMAN	P49238
GPR15	GPRF	GPRF_HUMAN	P49685
GPR17	GPRH	GPRH_HUMAN	Q13304
GPR18	GPRI	GPRI_HUMAN	Q14330
GPR20	GPRK	GPRK_HUMAN	Q99678
GPR21	GPRL	GPRL_HUMAN	Q99679
GPR22	GPRM	GPRM_HUMAN	Q99680
GPR24	GPRO	GPRO_HUMAN	Q99705
GPR25	GPRP	GPRP_HUMAN	O00155
GPR30	-	CML2_HUMAN	Q99527
GPR31	GPRV	GPRV_HUMAN	O00270
GPR38	-	GP38_HUMAN	O43193
GPR39	-	GP39_HUMAN	O43194
GPR40	-	GP40_HUMAN	O14842
GPR41	-	GP41_HUMAN	O14843
GPR42	-	GP42_HUMAN	O15529
GPR43	-	GP43_HUMAN	O15552
Platelet-activating receptor homologue	H963		O14626
HM74	HM74	HM74_HUMAN	P49019
KIAA0001	KI01	KI01_ HUMAN	Q15391

Leukotriene receptor (P2Y7)	-	P2Y7_HUMAN	Q15722
Mas-related receptor	-	MRG_HUMAN	P35410
Neurokinin I receptor	NK1R	NK1R_HUMAN	P25103
Nociceptin receptor	OPRX	OPRX_HUMAN	P41146
-	O15218		O15218
Delta-opioid receptor	OPRD	OPRD_HUMAN	P41143
Kappa-opioid receptor	OPRK	OPRK_HUMAN	P41145
Mu-opioid receptor	OPRM	OPRM_HUMAN	P35372
OGR1	OGR1	OGR1_HUMAN	Q15743
RDC1	RDC1	RDC1_ HUMAN	P25106
Somatostatin receptor 1	SSR1	SSR1_HUMAN	P30872
Somatostatin receptor 2	SSR2	SSR2_HUMAN	P30874
Somatostatin receptor 3	SSR3	SSR3_HUMAN	P32745
Somatostatin receptor 4	SSR4	SSR4_HUMAN	P31391
Somatostatin receptor 5	SSR5	SSR5_HUMAN	P35346

CLAIMS

1. A method of predicting a site for mutation of a first cellular receptor wherein the mutation alters the activity of the first cellular receptor, comprising the steps of:

(a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same receptor family;

(b) allocating the first cellular receptor to a receptor sub-family; and

(c) selecting an amino acid residue of the first cellular receptor for mutation, wherein the amino acid residue is analogous to a residue, the mutation of which is known to cause altered activity of a second cellular receptor, whereby a site for mutation of the first cellular receptor is predicted.

2. The method of claim 1 wherein said first and second cellular receptors are cell surface receptors.

3. The method of claim 1 wherein said first and second cellular receptors are G-protein coupled receptors.

4. The method of claim 3 wherein said first G protein-coupled receptor is an orphan receptor.

5. The method of claim 3 wherein step (b) is performed by phylogenetic profiling.

6. The method of claim 3 wherein the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of a human α_{1B} adrenoceptor.

7. The method of claim 3 wherein the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

8. The method of claim 3 wherein the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a close relative of the first G protein-coupled receptor.

9. The method of claim 3 wherein the second G protein-coupled receptor is in the same sub-family as the first G protein-coupled receptor.

10. The method of claim 1 wherein the mutation causes the receptor to become constitutively activated.

11. A method of obtaining a mutant of a first cellular receptor, wherein the mutant has altered activity as compared to wild type first cellular receptor, comprising the steps of:

(a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same family;

(b) allocating the first cellular receptor to a cellular receptor sub-family;

(c) selecting an amino acid residue of said first cellular receptor for mutation, wherein the selected amino acid residue is analogous to a residue of a second cellular receptor, the mutation of which is known to cause altered activity of said second cellular receptor;

(d) mutating the selected amino acid residue of the cellular receptor; and

(e) expressing the mutated cellular receptor in a cell.

12. The method of claim 11 wherein said first and second cellular receptors are cell surface receptors.

13. The method of claim 11 wherein said first and second cellular receptors are G-protein coupled receptors.

14. The method of claim 13 wherein the first G protein-coupled receptor is an orphan receptor.

15. The method of claim 11 wherein step (b) is performed by phylogenetic profiling.

16. The method of claim 13 wherein the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

17. The method of claim 13 wherein the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

18. The method of claim 11 wherein the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a close relative of the first cellular receptor.

19. The method of claim 11 wherein the second cellular receptor is in the same sub-family as the first cellular receptor.

20. The method of claim 11 wherein the mutation causes the receptor to become constitutively activated.

21. A method of predicting a site for mutation of a first cellular receptor wherein the mutation alters the activity of said first cellular receptor, said method comprising the steps of:

(a) performing a multiple sequence alignment of the cellular receptor with other cellular receptors in the same receptor family;

(b) allocating said cellular receptor to a receptor sub-family;

(c) determining whether mutant data are available for a member of the same sub-family as said cellular receptor; wherein if mutant data are available, then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue, the mutation of which is known to cause altered activity of said member of the same sub-family as the cellular receptor, whereby a site for mutation of said cellular receptor is predicted.

22. The method of claim 21 wherein said first cellular receptor is a cell surface receptor.

23. The method of claim 21 wherein said first cellular receptor is a G-protein coupled receptor.

24. The method of claim 23 wherein the first G protein-coupled receptor is an orphan receptor.

25. The method of claim 23 wherein the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

26. The method of claim 23 wherein the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

27. The method of claim 21 wherein step (b) is performed by phylogenetic profiling.

28. The method of claim 21 wherein the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as said first cellular receptor.

29. The method of claim 21 wherein the second cellular receptor is in the same sub-family as the first cellular receptor.

30. The method of claim 21 wherein the mutation causes the receptor to become constitutively activated.

31. A method of predicting a site for mutation of a first G protein-coupled receptor wherein the mutation alters the activity of the G protein-coupled receptor, comprising the steps of:

(a) performing a multiple sequence alignment of said first G protein-coupled receptor with other G protein-coupled receptors;

(b) allocating said first G protein-coupled receptor to a G protein-coupled receptor sub-family;

(c) determining whether mutant data are available for a member of the same sub-family as said first G protein-coupled receptor; wherein if mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a close relative of

the G protein-coupled receptor; and wherein if mutant data are not available then selecting an amino acid residue for mutation by identifying an amino acid residue selected from the group consisting of a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor and an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor; andwhereby a site for mutation of a G protein-coupled receptor is predicted.

32. The method of claim 31 wherein said first G protein-coupled receptor is an orphan receptor.

33. The method of claim 31 wherein step (b) is performed by phylogenetic profiling.

34. The method of claim 31 wherein the mutation causes the receptor to become constitutively activated.

35. A method of obtaining a mutant of a firstcellular receptor, wherein the mutant has altered activity as compared to wild type of said cellular receptor, comprising the steps of:

(a) performing a multiple sequence alignment of the first cellular receptor with other cellular receptors in the same family;

(b) allocating the cellular receptor to a cellular receptor sub-family;

(c) determining whether mutant data are available for a member of the same sub-family as said cellular receptor; wherein if mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a member of the same sub-family as said cellular receptor;

(d) mutating the selected amino acid residue of the cellular receptor; and

(e) expressing the mutated cellular receptor in a cell.

36. The method of claim 35 wherein said first and second cellular receptor is a cell surface receptor.

37. The method of claim 35 wherein said first cellular receptor is a G-protein coupled receptor.

38. The method of claim 37 wherein the first G protein-coupled receptor is an orphan receptor.

39. The method of claim 37 wherein the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

40. The method of claim 37 wherein the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

41. The method of claim 35 wherein step (b) is performed by phylogenetic profiling.

42. The method of claim 35 wherein the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a member of the same sub-family as the first cellular receptor.

43. The method of claim 35 wherein the second cellular receptor is in the same sub-family as the first cellular receptor.

44. The method of claim 35 wherein the mutation causes the receptor to become constitutively activated.

45. A method of obtaining a mutant of a first G protein-coupled receptor, wherein the mutant has altered activity as compared to wild type of said G protein-coupled receptor, comprising the steps of:

(a) performing a multiple sequence alignment of the first G protein-coupled receptor with other G protein-coupled receptors;

(b) allocating the G protein-coupled receptor to a G protein-coupled receptor sub-family;

(c) determining whether mutant data are available for a member of the same sub-family as the first G protein-coupled receptor; wherein if mutant data are available then selecting an amino acid residue for mutation, wherein the selected amino acid residue is analogous to a residue whose mutation is known to cause altered activity of a member of the same sub-family as said first G protein-coupled receptor; and wherein if mutant data are not available then selecting an amino acid residue for mutation by identifying an amino acid residue selected from the group consisting of a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor and an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human $_{1B}$ adrenoceptor;

(d) mutating the selected amino acid residue of the G protein-coupled receptor; and

(e) expressing the mutated G protein-coupled receptor in a cell.

46. The method of claim 45 wherein the first G protein-coupled receptor is an orphan receptor.

47. The method of claim 45 wherein step (b) is performed by phylogenetic profiling.

48. The method of claim 45 wherein the mutation causes the receptor to become constitutively activated.

49. The method of claim 45 wherein the amino acid residue selected for mutation is an amino acid residue that is 16 residues N-terminal to a conserved proline, wherein the conserved proline is analogous to P309 in transmembrane region 6 of the human α_{1B} adrenoceptor.

50. The method of claim 45 wherein the amino acid residue selected for mutation is a conserved asparagine residue analogous to N111 in transmembrane region 3 of a type 1A angiotensin II receptor.

51. The method of claim 45 wherein the amino acid selected for mutation is analogous to an amino acid known to cause altered activity in a close relative of the first cellular receptor.

52. The method of claim 45 wherein the second cellular receptor is in the same sub-family as the first cellular receptor.

53. A mutated GPR8 receptor comprising altered activity as compared to a wild type GPR8 receptor, wherein the GPR8 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 124 from aspartate to alanine, a mutation at amino acid 127 from asparagine to alanine and a mutation at amino acid 259 from threonine to glutamate.

54. A mutated GPR7 receptor comprising altered activity as compared to a wild type GPR7 receptor, wherein the GPR7 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 116 from aspartate to alanine, a mutation at amino acid 119 from asparagine to alanine and a mutation at amino acid 250 from threonine to glutamate.

55. A mutated GPR10 receptor comprising altered activity as compared to a wild type GPR10 receptor, wherein the GPR10 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 244 from tyrosine to glutamate and a mutation at amino acid 247 from valine to glutamate.

56. A mutated GPR1 receptor comprising altered activity as compared to a wild type GPR1 receptor, wherein the GPR1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 176 from arginine to alanine, a mutation at amino acid 245 from phenylalanine to glutamate, and a mutation at amino acid 120 from asparagine to alanine.

57. A mutated GPR17 receptor comprising altered activity as compared to a wild type GPR17 receptor, wherein the GPR17 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 114 from asparagine to alanine and a mutation at amino acid 234 from valine to glutamate.

58. A mutated GPR4 receptor comprising altered activity as compared to a wild type GPR4 receptor, wherein the GPR4 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 100 from asparagine to alanine and a mutation at amino acid 223 from lysine to glutamate.

59. A mutated GPR15 receptor comprising altered activity as compared to a wild type GPR15 receptor, wherein the GPR15 receptor comprises a mutation selected from the

group consisting of a mutation at amino acid 116 from asparagine to alanine and a mutation at amino acid 240 from isoleucine to glutamate.

60. A mutated GPR20 receptor comprising altered activity as compared to a wild type GPR20 receptor, wherein the GPR20 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 133 from asparagine to alanine and a mutation at amino acid 240 from methionine to glutamate.

61. A mutated HM74 receptor comprising altered activity as compared to a wild type HM74 receptor, wherein the HM74 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 110 from asparagine to alanine and a mutation at amino acid 230 from isoleucine to glutamate.

62. A mutated OGR1 receptor comprising altered activity as compared to a wild type OGR1 receptor, wherein the OGR1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 104 from asparagine to alanine and a mutation at amino acid 227 from glutamine to glutamate.

63. A mutated EBI2 receptor comprising altered activity as compared to a wild type EBI2 receptor, wherein the EBI2 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 114 from asparagine to alanine and a mutation at amino acid 243 from leucine to glutamate.

64. A mutated BONZO receptor comprising altered activity as compared to a wild type BONZO receptor, wherein the BONZO receptor comprises a mutation selected from the group consisting of a mutation at amino acid 112 from asparagine to alanine and a mutation at amino acid 230 from leucine to glutamate.

65. A mutated RDC1 receptor comprising altered activity as compared to a wild type RDC1 receptor, wherein the RDC1 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 127 from asparagine to alanine and a mutation at amino acid 251 from arginine to glutamate.

66. A mutated O15218 receptor comprising altered activity as compared to a wild type O15218 receptor, wherein the O15218 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 136 from asparagine to alanine and a mutation at amino acid 257 from cysteine to glutamate.

67. A mutated H963 receptor comprising altered activity as compared to a wild type H963 receptor, wherein the H963 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 97 from asparagine to alanine and a mutation at amino acid 222 from leucine to glutamate.

68. A mutated GPR30 receptor comprising altered activity as compared to a wild type GPR30 receptor, wherein the GPR30 receptor comprises a mutation selected from the group consisting of a mutation at amino acid 140 from asparagine to alanine and a mutation at amino acid 258 from leucine to glutamate.

69. A mutated GPR2 receptor comprising altered activity as compared to a wild type GPR2 receptor, wherein the GPR2 receptor comprises a mutation at amino acid 238 from leucine to glutamate.

70. A mutated GPR5 receptor comprising altered activity as compared to a wild type GPR5 receptor, wherein the GPR5 receptor comprises a mutation at amino acid 224 from valine to glutamate.

71. A mutated GPR13 receptor comprising altered activity as compared to a wild type GPR13 receptor, wherein the GPR13 receptor comprises a mutation at amino acid 230 from isoleucine to glutamate.

72. A mutated GPR18 receptor comprising altered activity as compared to a wild type GPR18 receptor, wherein the GPR18 receptor comprises a mutation at amino acid 231 from isoleucine to glutamate.

73. A mutated GPR21 receptor comprising altered activity as compared to a wild type GPR21 receptor, wherein the GPR21 receptor comprises a mutation at amino acid 251 from alanine to glutamate.

74. A mutated GPR22 receptor comprising altered activity as compared to a wild type GPR22 receptor, wherein the GPR22 receptor comprises a mutation at amino acid 312 from phenylalanine to glutamate.

75. A mutated GPR25 receptor comprising altered activity as compared to a wild type GPR25 receptor, wherein the GPR25 receptor comprises a mutation at amino acid 242 from leucine to glutamate.

76. A mutated GPR31 receptor comprising altered activity as compared to a wild type GPR31 receptor, wherein the GPR31 receptor comprises a mutation at amino acid 221 from glutamine to glutamate.

77. A mutated GPR38 receptor comprising altered activity as compared to a wild type GPR38 receptor, wherein the GPR38 receptor comprises a mutation at amino acid 297 from valine to glutamate.

78. A mutated GPR39 receptor comprising altered activity as compared to a wild type GPR39 receptor, wherein the GPR39 receptor comprises a mutation at amino acid 282 from isoleucine to glutamate.

79. A mutated GPR40 receptor comprising altered activity as compared to a wild type GPR40 receptor, wherein the GPR40 receptor comprises a mutation at amino acid 223 from alanine to glutamate.

80. A mutated GPR41 receptor comprising altered activity as compared to a wild type GPR41 receptor, wherein the GPR41 receptor comprises a mutation at amino acid 224 from alanine to glutamate.

81. A mutated GPR42 receptor comprising altered activity as compared to a wild type GPR42 receptor, wherein the GPR42 receptor comprises a mutation at amino acid 224 from alanine to glutamate.

82. A mutated GPR43 receptor comprising altered activity as compared to a wild type GPR43 receptor, wherein the GPR43 receptor comprises a mutation at amino acid 221 from valine to glutamate.

83. A mutated MGR receptor comprising altered activity as compared to a wild type MGR receptor, wherein the MGR receptor comprises a mutation at amino acid 263 from tyrosine to glutamate.

84. A method of identifying a compound which modulates the activity of a receptor as claimed in any one of claims 53-83, the method comprising

a) contacting a candidate compound with said receptor, and

b) determining activity of said receptor in the presence of said compound, wherein a difference in receptor activity in the presence and absence of said candidate compound is indicative of compound modulation.

85. The method of claim 84, wherein said compound is further determined to be an inverse agonist, partial agonist or an agonist of said receptor activity.

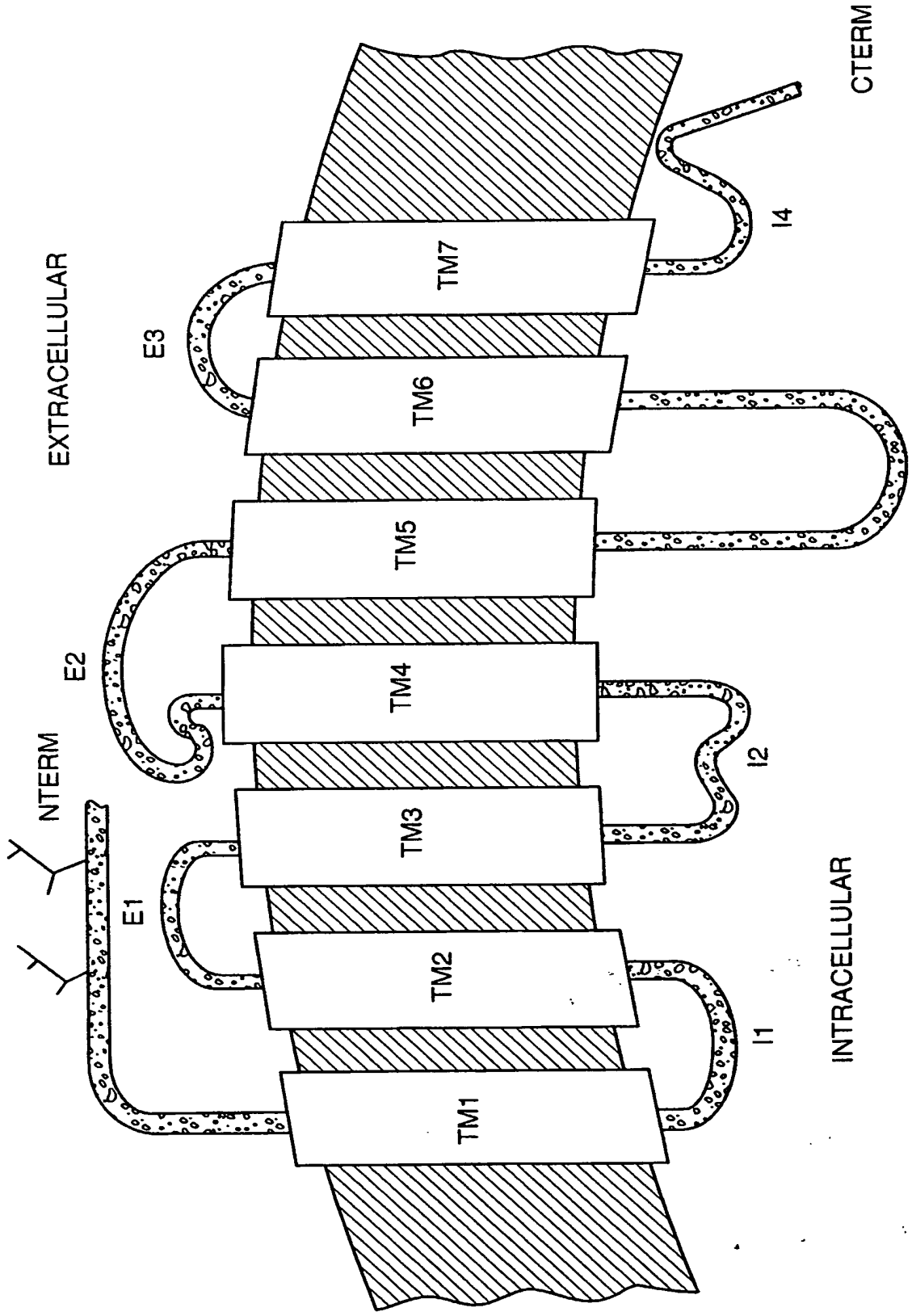

EXTRACELLULAR
INTRACELLULAR
NTERM
CTERM
TM1
TM2
TM3
TM4
TM5
TM6
TM7
E1
E2
E3
I1
I2
I3
I4

FIG. 1

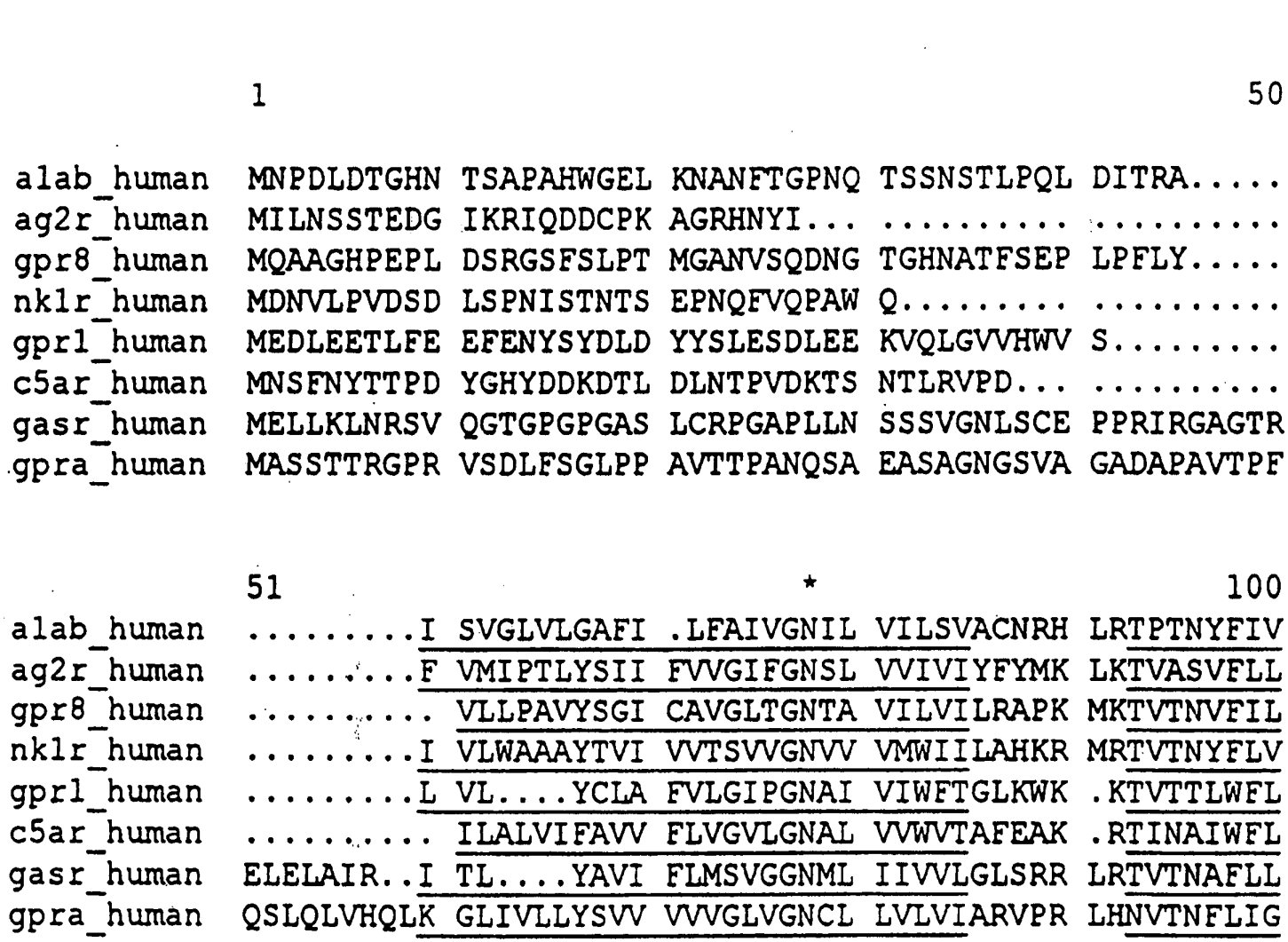


FIG. 2A

FIG. 2A
FIG. 2B
FIG. 2C
FIG. 2D
FIG. 2E
FIG. 2F

FIG. 2

```
              1012    *                                     *            150
alab_human    NLAMADLLLS FTVLPFSAAL EVLGY.WVLG RIFCDIWAAV DVLCCTASIL
ag2r_human    NLALADLCFL LT.LPLWAVY TAMEYRWPFG NYLCKIASAS VSFNLYASVF
gpr8_human    NLAVADGLFT LV.LPVNIAE HLLQY.WPFG ELLCKLVLAV DHYNIFSSIY
nk1r_human    NLAFAEASMA AFNTVVNFTY AV.HNEWYYG LFYCKFHNFF PIAAVFASIY
gpr1_human    NLAIADFIFL LF.LPLYISY VAMNFHWPFG IWLCKANSFT AQLNMFASVF
c5ar_human    NLAVADFLSC LA.LPILFTS IVQHHHWPFG GAACSILPSL ILLNMYASIL
gasr_human    SLAVSDLLLA VACMPFTLLP NLMGT.FIFG TVICKAVSYL MGVSVSVSTL
gpra_human    NLALSDVLMC TACVPLTLAY AFEPRGWVFG GGLCHLVFFL QPVTVYVSVF
                      *                                     *

              151     *                                    *        200
alab_human    SLCAISIDRY IGVRYSLQYP ..TLVTRRKA ILALLSVWVL STVISIGPLL
ag2r_human    LLTCLSIDRY LAIVHPMKSR ..LRRTMLVA KVTCIIIWLL AGLASLPAII
gpr8_human    FLAVMSVDRY LVVLATVRSR HMPWRTYRGA KVASLCVWLG VTVLVLPFFS
nk1r_human    SMTAVAFDRY MAIIHPLQPR ..LSAT..AT KVVICVIWVL ALLLAFPQGY
gpr1_human    FLTVISLDHY IHLIHPVLSH ..RHRTLKNS LIVIIFIWLL ASLIGGPALY
c5ar_human    LLATISADRF LLVFKPIWCQ NFRGAGL..A WIACAVAWGL ALLLTIPSFL
gasr_human    SLVAIALERY SAICRPLQAR ..VWQTRSHA ARVIVATWLL SGLLMVPYPV
gpra_human    TLTTIAVDRY VVLVHPLR.. ...RASRCAS AYAVLAIWAL SAVLALPPAV
                      *                                    *
```

FIG. 2B

```
                201                *                                  250
a1ab_human   ..GWKEPAPN DDKEC..... GVTEEPFYAL FSSLGSFYIP LAVILVMY..
ag2r_human   HRNVFFIENT NITVCAFHYE SQNSTLPI.. ..GLGLTKNI LGFLFPFLII
gpr8_human   FAGVYSNEL. QVPSCGLSFP .WPER...VW FKASRVYTLV LGFVLPVCTI
nk1r_human   YSTT..ETMP SRVVCMIEWP EHPNK...IY EKVYHICVTV LIYFLPLLVI
gpr1_human   FRDT..VEFN NHTLCYNNFQ KHDPDLTLIR HHVLTWVKFI IGYLFPLLTM
c5ar_human   YRVVREEYFP PKVLCGVDY. SHDKR....R ERAVAIVRLV LGFLWPLLTL
gasr_human   YTVV.QPVGP RVLQCVHRWP SARVRQTWSV LLLLLLFFIP GVVMAVAYGL
gpra_human   HTYHVELKPH DVRLCEEFWG SQERQ..... RQLYAWGLLL VTYLLPLLVI
                                   *

                251                                                300
a1ab_human   .CRVYIVAKR TTKNLEAGVM KEMSNSKELT LRIHSKN... .....FHEDT
ag2r_human   LTSYTLIWKA LKKAYEIQKN KPRNDD.... .......... ..........
gpr8_human   CVLYTDLLRR LRAVRL.... .RSGAK.... .......... ..........
nk1r_human   GYAYTVVGIT LWASEIPGDS SDRYHE.... .......... ..........
gpr1_human   SICYLCLIFK VKKRTVLISS RH........ .......... ..........
c5ar_human   TICYTFILLR TWSRRATRST KT........ .......... ..........
gasr_human   ISRELYLGLR FDGDSDSDSQ SRVRNQGGLP GAVHQNGRCR PETGAVGEDS
gpra_human   LLSYVRVSVK LRNRVVPGCV TQSQAD.... .......... ..........
```

FIG. 2C

```
                                                                   1      350
            301
alab_human  .......... ...LSSTKAK GHNPRSSIAV KLFKFSREKK AAKTLGIVVG
ag2r_human  .......... .......... .......... .......... IFKIIMAIVL
gpr8_human  .......... .......... .......... ..ALGKARRK VTVLVLVVLA
nk1r_human  .......... .......... .......... ..QVSA.KRK VVKMMIVVVC
gpr1_human  .......... .......... .......... .......... .FWTILVVVV
c5ar_human  .......... .......... .......... .......... .LKVVVAVVA
gasr_human  DGCYVQLPRS RPALELTALT APGPGSGSRP TQAKLLAKKR VVRMLLVIVV
gpra_human  .......... .......... .......... ..WDRARRRR TFCLLVVVVV

            351   *                                             400
alab_human  MFILCWLPFF IALPLGSLFS TLKPPD.... ....AVFKVV FWLGYFNSCL
ag2r_human  FFFFSWIPHQ IFTFLDVLIQ LGIIRDCRIA DIVDTAMPIT ICIAYFNNCL
gpr8_human  VCLLCWTPFH LASVVALTTD L....PQTPL VISMSYV..I TSLTYANSCL
nk1r_human  TFAICWLPFH IFFLLPYI.. ...NPDLYLK KFIQQVYLAI MWLAMSSTMY
gpr1_human  AFVVCWTPYH LFSIWELTIH H....NSYSH HVMQAGIPLS TGLAFLNSCL
c5ar_human  SFFIFWLPYQ VTGIMMSFLE P....SSPTF LLLNKLDSLC VSFAYINCCI
gasr_human  LFFLCWLPVY SANTWRAF.. ...DGPGAHR ALSGAPISFI HLLSYASACV
gpra_human  VFAVCWLPLH VFNLLRDL.. ...DPHAIDP YAFGLVQLLC HWLAMSSACY
                  *
```

FIG. 2D

```
          401 *                                                    450
a1ab_human  NPIIYPCSSK EFKRAFVRIL GCQCRGRRRR RRRRRRLGGC AYTYRPWTRG
ag2r_human  NPLFYGFLGK KFKRYFLQLL KYIPP..... .......... ..........
gpr8_human  NPFLYAFLDD NFRKNFRSIL RC........ .......... ..........
nk1r_human  NPIIYCCLND RFRLGFKHAF RC.CPFISAG DYEGLEMKST RYLQ...TQG
gpr1_human  NPILYVLISK KFQARFRSSV AEI....... .......... ..........
c5ar_human  NPIIYVVAGQ GFQGRLRKSL PSL....... .......... ..........
gasr_human  NPLVYCFMHR RFRQACLETC ARCCP..... .......... ..........
gpra_human  NPFIYAWLHD SFREELRKLL VAWPR..... .......... ..........
             *

            451                                                    500
a1ab_human  GSLERSQSRK DSLDDSGSCL SGSQRTLPSA SPSPGYLGRG APPPVELCAF
ag2r_human  .......... .......... .......... .......... ..........
gpr8_human  .......... .......... .......... .......... ..........
nk1r_human  SVYKVSR... ..LETTISTV VGAHEEEPED GP......KA TPSSLDLTS.
gpr1_human  .......... .......... .......... .......... ..........
c5ar_human  .......... .......... .......... .......... ..........
gasr_human  .......... .......... .......... .......... ..........
gpra_human  .......... .......... .......... .......... ..........
```

FIG. 2E

```
                                                                  550
                501                                               |
alab_human      PEWKAPGALL SLPAPEPPGR RGRHDSGPLF TFKLLTEPES PGTDGGASNG
ag2r_human      ..........  ..........  ..........  ..........K AKSHSNLSTK
gpr8_human      ..........  ..........  ..........  .......... ..........
nk1r_human      ..........  ..........NC SSRSDS....  ..KTMTESFS ..........
gpr1_human      ..........  ..........  ..........  .......... ...LKYTLWE
c5ar_human      ..........  ..........  ..........  .......... ...LRNVLTE
gasr_human      ..........  ..........  ..........  ..........R PPRARPRALP
gpra_human      ..........  ..........  ..........  ..........K IAPHGQNMT.

                551                                               576
alab_human      GCEAAADVAN GQPGFKSNMP LAPGQF
ag2r_human      MSTLSYRPSD NVSSSTKKPA PCFEVE
gpr8_human      ..........  ..........  ......
nk1r_human      ..........  ....FSSNV.  ....LS
gpr1_human      VSCSGTVSEQ LRNSETKNLC LLETAQ
c5ar_human      ESVV.RESKS FTRSTVDTMA QKTQAV
gasr_human      DEDPPTPSIA SLSRLSYTTI STLGPG
gpra_human      ..........  ..........  .VSVVI
```

FIG. 2F

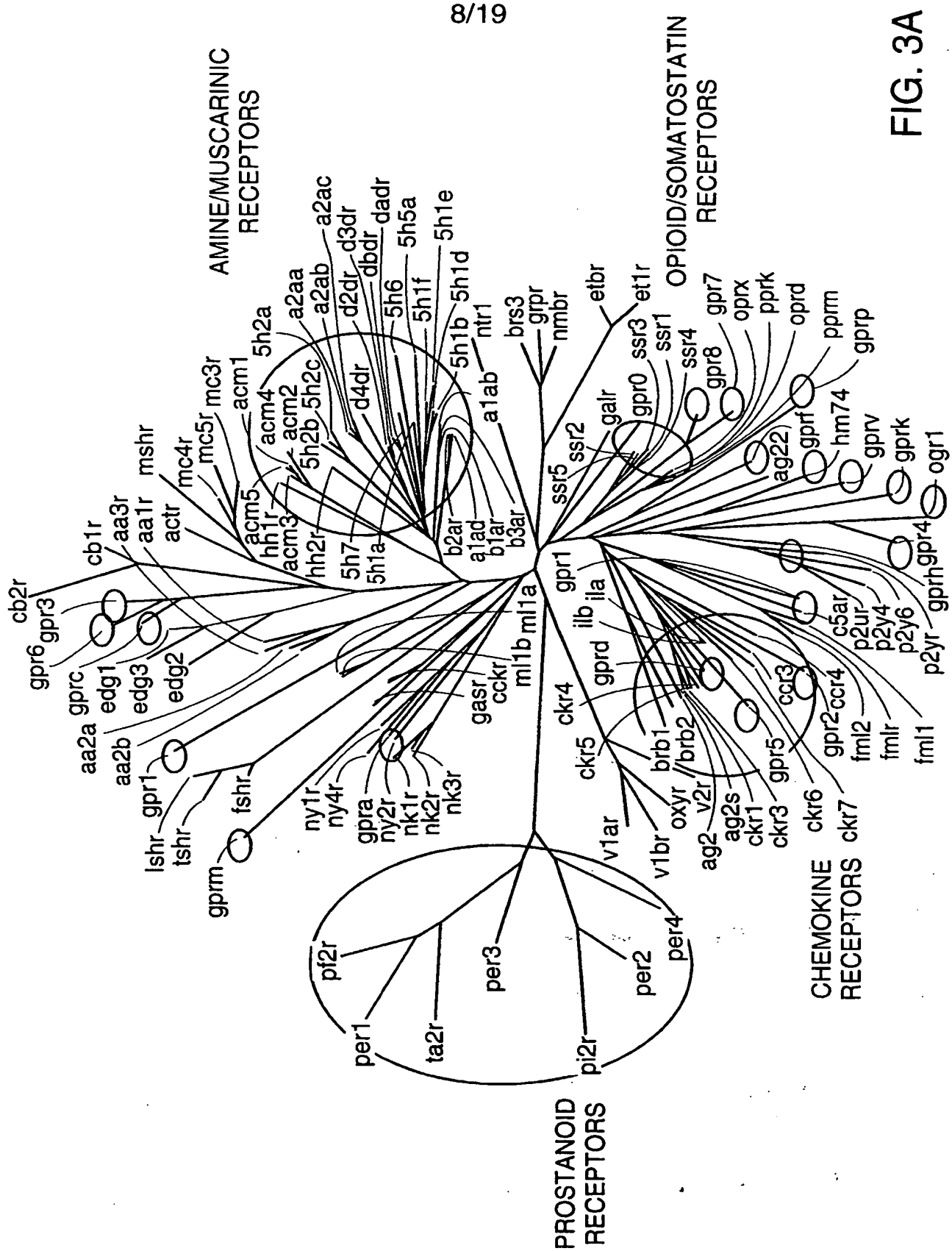

AMINE/MUSCARINIC RECEPTORS
OPIOID/SOMATOSTATIN RECEPTORS
CHEMOKINE RECEPTORS
PROSTANOID RECEPTORS
cb2r
gpr3
gpr6
gprc
edg1
edg3
edg2
aa2a
aa2b
gpr1
lshr
tshr
gprm
fshr
cb1r
aa3r
aa1r
actr
mshr
mc4r
mc5r
mc3r
acm1
acm4
acm2
acm5
hh1r
acm3
hh2r
5h2a
a2aa
a2ab
a2ac
d2dr
d3dr
dbdr
dadr
5h2b
5h2c
d4dr
5h6
5h5a
5h1f
5h1e
5h1b
5h1d
ntr1
5h7
5h1a
b2ar
a1ad
b1ar
b3ar
a1ab
brs3
grpr
nmbr
etbr
et1r
ssr5
ssr2
galr
gpr0
ssr3
ssr1
ssr4
gpr8
gpr7
oprx
pprk
oprd
pprm
gprp
ag22
gprr
hm74
gprv
gprk
ogr1
gpr4
gprh
p2yr
p2y6
p2y4
p2ur
c5ar
gpr1
ml1a
ml1b
cckr
gasr
ny1r
ny4r
gpra
ny2r
nk1r
nk2r
nk3r
ilb
ila
gprd
ckr4
ckr5
brb1
brb2
v1ar
v1br
oxyr
v2r
ag2
ag2s
ckr1
ckr3
gpr5
ccr3
ccr4
gpr2
fml2
fmlr
fml1
ckr6
ckr7
pf2r
per1
ta2r
per3
pi2r
per2
per4

FIG. 3A

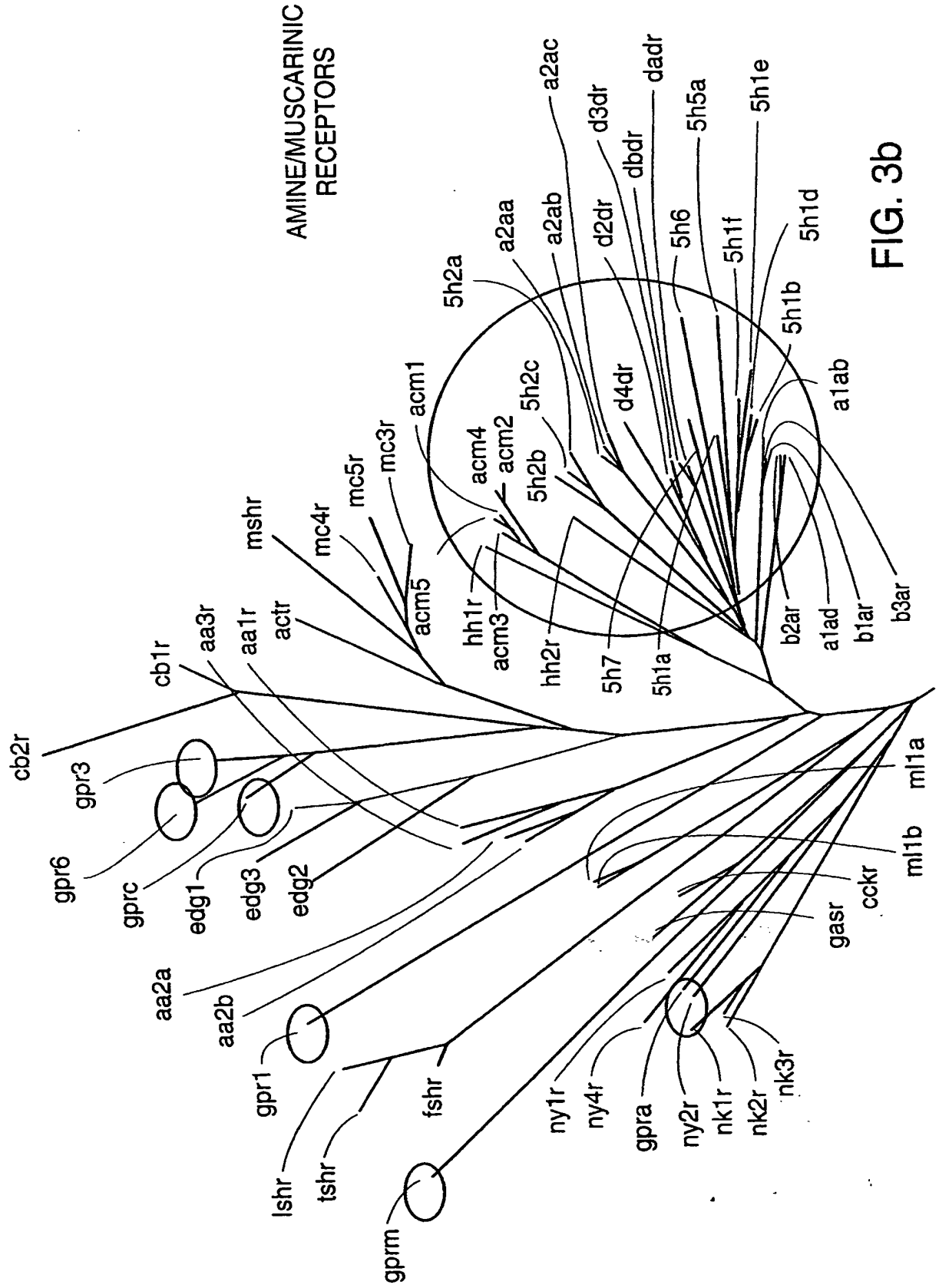

AMINE/MUSCARINIC RECEPTORS
cb2r
gpr3
gpr6
gprc
edg1
edg3
edg2
cb1r
aa3r
aa1r
actr
mshr
mc4r
mc5r
mc3r
acm1
acm5
hh1r
acm3
acm4
acm2
5h2a
5h2c
5h2b
hh2r
5h7
5h1a
a2aa
a2ac
a2ab
d2dr
d3dr
d4dr
dbdr
dadr
5h6
5h5a
5h1f
5h1e
5h1b
5h1d
a1ab
b2ar
a1ad
b1ar
b3ar
ml1a
ml1b
cckr
gasr
aa2a
aa2b
gpr1
lshr
tshr
fshr
gprm
ny1r
ny4r
gpra
ny2r
nk1r
nk2r
nk3r

FIG. 3b

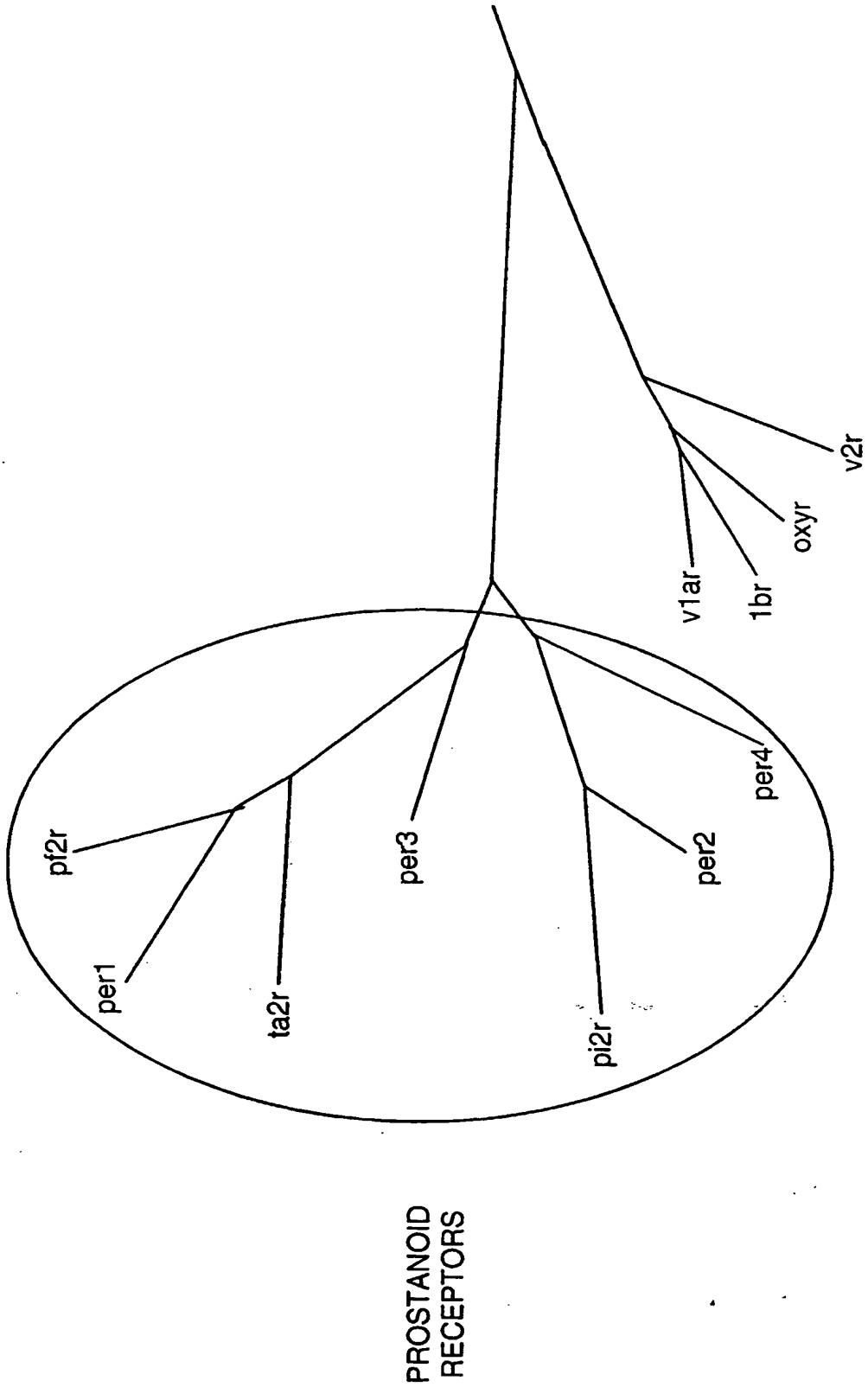

v2r
oxyr
1br
v1ar
per4
per2
per3
pi2r
ta2r
per1
pf2r
PROSTANOID
RECEPTORS

FIG. 3c

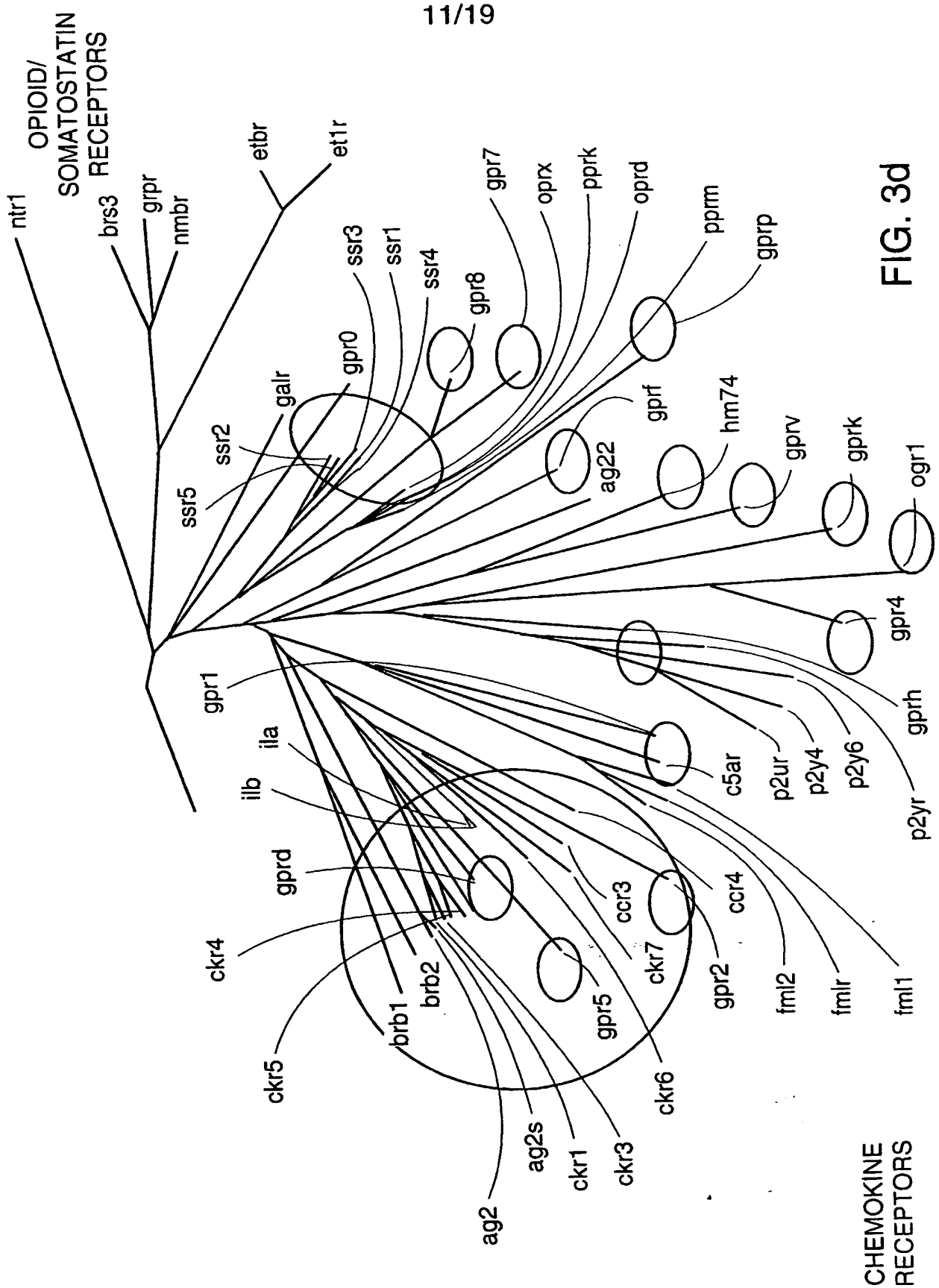

OPIOID/
SOMATOSTATIN
RECEPTORS
ntr1
brs3
grpr
nmbr
etbr
et1r
ssr5
ssr2
galr
gpr0
ssr3
ssr1
ssr4
gpr8
gpr7
oprx
pprk
oprd
pprm
gprp
ag22
gprf
hm74
gprv
gprk
ogr1
gpr4
gprh
p2yr
p2y6
p2y4
p2ur
c5ar
ccr4
fml2
fmlr
fml1
gpr2
ckr7
ccr3
gpr5
ckr6
ckr3
ckr1
ag2s
ag2
ckr5
brb1
brb2
ckr4
gprd
ilb
ila
gpr1
CHEMOKINE
RECEPTORS

FIG. 3d

```
*                    *
gpr8_human   113-GELLCKLVLAVDHYNIFSSIYFLAVMSVDRYLVVLATVRS-153
                          A
oprd_human   117-GELLCKAVLSIDYYNMFTSIFTLTMMSVDRYIAVCHPVKA-157
oprm_human   138-GTILCKIVISIDYYNMFTSIFTLCTMSVDRYIAVCHPVKA-178
oprk_human   127-GDVLCKIVISIDYYNMFTSIFTLTMMSVDRYIAVCHPVKA-167
oprx_human   119-GNALCKTVIAIDYYNMFTSTFTLTAMSVDRYVAICHPIRA-159
ssr1_human   126-GALLCRLVLSVDAVNMFTSIYCLTVLSVDRYVAVVHPIKA-166
ssr2_human   111-GKAICRVVMTVDGINQFTSIFCLTVMSIDRYLAVVHPIKS-151
ssr3_human   112-GSLMCRLVMAVDGINQFTSIFCLTVMSVDRYLAVVHPTRS-152
ssr4_human   115-GSVLCRAVLSVDGLNMFTSVFCLTVLSVDRYVAVVHPLRA-155
ssr5_human   108-GPVLCRLVMTLDGVNQFTSVFCLTVMSVDRYLAVVHPLSS-148
```

FIG. 4

```
                                                    *
gpr8_human   255-RRKVTVLVLVVLAVCLLCWTPFHLAS-280
                                                    E
alab_human   289-EKKAAKTLGIVVGMFILCWLPFFIAL-314
```

FIG. 5A

```
                                                    *
gpr8_human   117-CKLVLAVDHYNIFSSIYFLAVMSVDRY-143
                                                    A
ag2r_human   101-CKIASASVSFNLYASVFLLTCLSIDRY-127
```

FIG. 5B

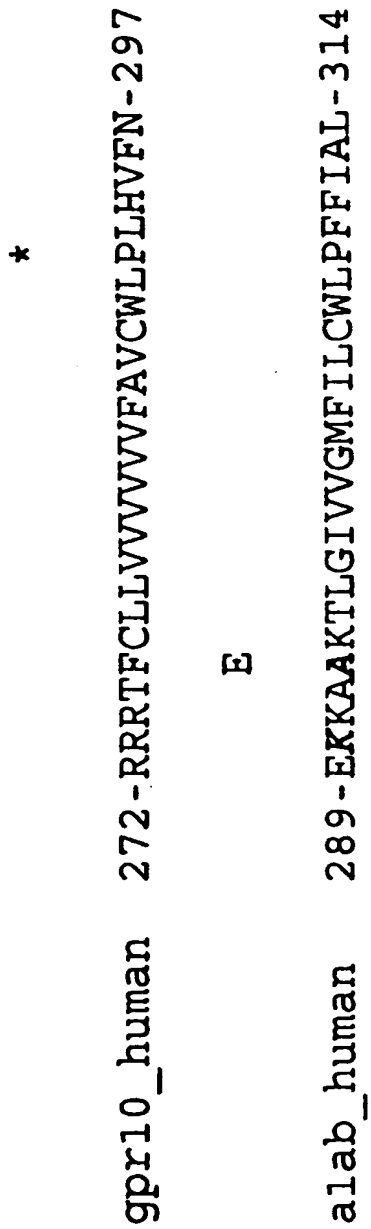

*
gpr10_human 272-RRRTFCLLVVVVVVFAVCWLPLHVFN-297
E
alab_human 289-EKKAAKTLGIVVGMFILCWLPFFIAL-314

FIG. 6A

*
gpr10_human 236-PLLVILLLSYVRVSVKL-251
E
nklr_human 208-PLLVIGYAYTVVGITL-223

FIG. 6B

```
                                        *
gpr1_human    234-QKRVFRMSLLIISTFLLCWTPYHLFS-266

                        E

a1ab_human    289-EKKAAKTLGIVVGMFILCWLPFFIAL-314
```

FIG. 7A

```
                *                       *
gpr1_human    110-CKANSFTAQLNMFASVFFLTVISLDHY-136

                        A

ag2r_human    101-CKIASASVSFNLYASVFLLTCLSIDRY-127
```

FIG. 7B

```
                                  *
gpr1_human   162-WLLASLIGGPALYFRDTVE..FNNHTLC-187
                                  A
c5ar_human   170-WGLALLLTIPSFLYRVVREEYFPPKVLC-188
fml1_human   150-WILALVLTLPVFLFLTTVTIP.NGDTYC-176
p2y7_human   142-WVLSFLLATPVLAYRTVVPWKTNMSL.C-168
```

FIG. 7C

```
                                                    *
gpr17_human  230-KTKAVRMIAIVLAIFLVCFVPYHVVNR-255

                                                    E

a1ab_human   289-EKKAAKTLGIVVGMFILCWLPFFIAL-314
```

FIG. 8A

```
                                                    *

                                                    *
gpr17_human  104-CRLTGFLFYLNMYASIYFLTCISADRF-130

                                                    A

ag2r_human   101-CKIASASVSFNLYASVFLLTCLSIDRY-127
```

FIG. 8B

gpr21_human 247-DKRYAMVLFRITSVFYILWLPYIIYF-272

E

alab_human 289-EKKAAKTLGIVVGMFILCWLPFFIAL-314

FIG. 9

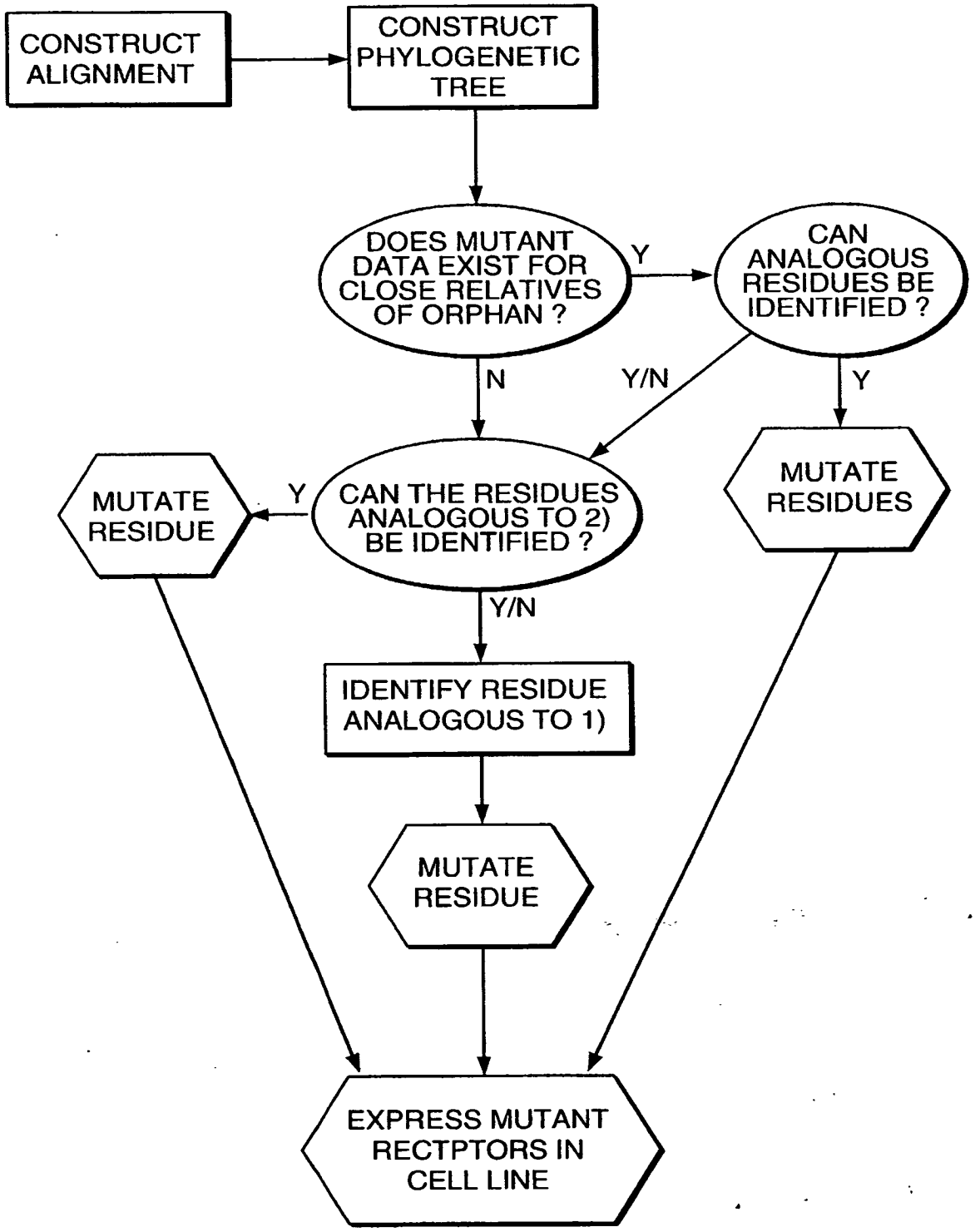

CONSTRUCT ALIGNMENT
CONSTRUCT PHYLOGENETIC TREE
DOES MUTANT DATA EXIST FOR CLOSE RELATIVES OF ORPHAN ?
Y
CAN ANALOGOUS RESIDUES BE IDENTIFIED ?
N
Y/N
Y
MUTATE RESIDUE
Y
CAN THE RESIDUES ANALOGOUS TO 2) BE IDENTIFIED ?
MUTATE RESIDUES
Y/N
IDENTIFY RESIDUE ANALOGOUS TO 1)
MUTATE RESIDUE
EXPRESS MUTANT RECTPTORS IN CELL LINE

FIG. 10

IN THE INTERNATIONAL BUREAU AS RECEIVING OFFICE

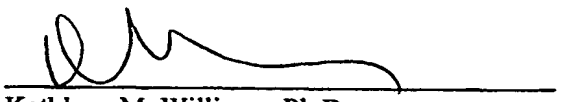
International Application No.	:	PCT/IB00/01407
International Filing Date	:	02 October 2000
Earliest Priority Date	:	29 September 1999
Applicant(s)	:	Cambridge Drug Discovery, Ltd., et al
Title	:	Method of Predicting Mutations
Authorized Officer	:	Agnes Wittmann-Regis

International Bureau of WIPO
PCT Receiving Office Section
1211 Geneva 20, Switzerland

RESPONSE TO THE INVITATION TO CORRECT PRIORITY CLAIM IN THE INTERNATIONAL APPLICATION

In response to the Invitation to Correct Priority Claim mailed in the above-referenced international application, Applicant mailed the above Application on 29 September 2000 with intent to claim priority to U.K. Application No. 9922986.6 dated 29 September 1999. Applicant acknowledges that the International Bureau does not recognize the mailing date of Application to be the date received by the Office, unlike the U.S. Receiving Office.

Respectfully submitted,



Kathleen M. Williams, Ph.D.
Common Representative
Palmer & Dodge, LLP
One Beacon Street
Boston, Massachusetts 02109
Telephone: 617-573-0451
Facsimile: 617-227-4420

DOCKET NO.: 4219/1370
DATE: 24 January 2001